Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Single-Molecule FRET Studies of Protein Function and Conformational Dynamics

From DNA Nanotechnology to Viral and Bacterial Infections

Kira Bartnik

aus

Mannheim, Deutschland

2020

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Don C. Lamb, Ph.D., betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 14.09.2020

Kira Bartnik

Dissertation eingereicht am:	22.07.2020
1. Gutachter	Prof. Don C. Lamb, Ph.D.
2. Gutachter	Prof. Dr. Christoph Bräuchle
Mündliche Prüfung am:	04.09.2020

Abstract

Single-Molecule FRET Studies of Protein Function and Conformational Dynamics - From DNA Nanotechnology to Viral and Bacterial Infections

by Kira BARTNIK

All fundamental cellular processes are governed by dynamic interactions of nano-sized biomolecules such as proteins and DNA. A technique well suited to resolve inter- and intramolecular distances is Förster Resonance Energy Transfer (FRET). By combining FRET with single-molecule spectroscopy, heterogeneities in the system can be detected and dynamic information becomes accessible. In this thesis, single-molecule FRET (smFRET) experiments were performed using either the small observation volume of a confocal microscope or the evanescent field of a total internal reflection (TIRF) microscope to get insights into the enzymatic activity of three proteins. To study the end-joining of DNA double strands by the T4 DNA ligase, a DNA origami platform was designed to control the stoichiometry and spatial arrangement of the interaction partners, thereby increasing their local concentration. The ligation process could be followed in real time and the same pair of DNA strands repeatedly ligated and cut, which highlights the applicability of the DNA origami platform to a wide range of multimolecular interaction studies on the single-molecule level. In a second project, the underlying catalytic mechanism of NSP2, a rotavirus protein required for genome replication and virus assembly, was investigated by comparing the RNA unwinding activity of full-length and C-terminally truncated mutants. Although the C-terminal region was less efficient in destabilizing the secondary structure of RNA, it proved to be essential for RNA release via charge repulsion and is thus a prerequisite for a full cycle of chaperone activity. In the last study, the conformational states of bacterial adhesion protein SdrG were studied. This protein initiates nosocomial infections through its stable attachment to human fibrinogen. By combining FRET-derived distances and molecular dynamics (MD) simulations, the structure of the dynamic apoprotein could be modeled, providing information inaccessible to other methods. The diverse topics researched in this thesis - ranging from DNA nanotechnology to viral and bacterial infections - emphasize the multifaceted capabilities of smFRET.

Contents

Chapter 1

Introduction

In fluorescence experiments using Förster Resonance Energy Transfer (FRET), the energy transferred between two spectrally different donor and acceptor fluorophores directly reports on their proximity. As a "spectroscopic ruler" that is sensitive to distances in the range of 1-10 nm (L Stryer, 1967), FRET is well suited to study biomolecules such as nucleic acids or proteins at the nanometer scale, and has become a widely used tool in biological sciences (Algar et al., 2019; Shrestha et al., 2015). The growing diversity of available FRET methods to tackle a plethora of research questions is reflected in the number of FRET-related publications, which has increased significantly over the last twenty years from less than 30 in 1989 to close to 1900 in 2019.[1] A major breakthrough in the field has been the realization of experiments with single-molecule sensitivity, pioneered by Taekip Ha and co-workers almost 25 years ago (Ha et al., 1996). By measuring one FRET pair at a time, heterogeneities can be resolved and dynamic transitions between different states observed. Single-molecule FRET (smFRET) has been used extensively ever since to investigate the conformational dynamics and interactions of individual biomolecules (Lerner et al., 2018) and is the main method utilized in this thesis to examine the enzymatic function of different proteins.

In its simplest form, FRET can be used as a readout to assess whether or not the acceptor and donor fluorophores are close together, for example, due to binding and unbinding of interaction partners. Two of the three projects presented in this work (schematically illustrated in Figure 1.1) are built on this qualitative aspect of FRET. To investigate the end-joining of two DNA double strands by the T4 DNA ligase - a model reaction to mimic the ligation step in DNA double-strand break repair - one DNA strand was labeled with a donor, and the other with an acceptor fluorophore. Before ligation, the labels on the individual DNA strands were far apart, while ligation connected the strands and brought the FRET pair in close proximity leading to an efficient energy transfer. To locally increase the concentration of the repair substrates, the DNA strands were attached to a custom-designed DNA origami platform. Using this experimental approach, transient binding of the DNA double strands could be investigated before enzyme addition and the transition from the free to the ligated state directly observed.
When both FRET dyes are on the same biomolecule, structural changes can be probed. In the second project presented in this thesis, donor and acceptor fluorophores were placed on the two termini of an RNA stem-loop to examine its conformational states. In its folded state, the FRET dyes were close together and the energy transfer efficient, while unfolding of the RNA pushed the inter-dye distance beyond the FRET-sensitive limit. Unfolding was triggered by the viral RNA chaperone protein NSP2 and a comparative study with C-terminally truncated mutants revealed the importance of the C-terminal region for an efficient RNA release via charge repulsion required for a full cycle of chaperone activity.
The third project of this thesis takes full advantage of the quantitative potential of smFRET. Placing not only one, but a network of FRET pairs on the bacterial adhesion protein SdrG made it possible to investigate its conformational changes in the absence and presence of the target peptide, human Fgβ. Accurate inter-dye distances were extracted from the data with photon distribution analysis (PDA) (Antonik et al., 2006). By comparing the experimental results to molecular dynamics (MD) simulations, we could propose a structural model of the dynamic apoprotein. The ability to simultaneously provide information on structure and dynamics makes FRET clearly stand out from other single-molecule techniques.

Following a short introduction on fluorescence, its fundamental properties, organic fluorophores and the concept of FRET in chapter 2, two experimental methods for the observation of single molecules are presented in chapter 3: confocal and total internal fluorescence (TIRF) microscopy, which were used in this work to measure freely diffusing and surface-immobilized molecules, respectively. Experimental details and analysis procedures are discussed with a focus on the calculation of accurate FRET efficiencies. Chapter 4 provides an

[1] numbers taken from *Web of Science*, searching for the topic 'FRET'; https://apps.webofknowledge.com (accessed June 5, 2020)

overview of the DNA ligation project using the DNA origami platform to increase the local concentration of the interaction partners. Chapter 5 summarizes the findings from the comparative study with C-terminally truncated and full-length NSP2 chaperone mutants and provides a model to understand the binding of NSP2 to RNA and the role of its C-terminal region in detail. In chapter 6, the binding of the bacterial adhesion protein SdrG to its target peptide Fgβ is investigated, which involves a conformational switch. SmFRET data are combined with molecular dynamics simulations to quantitatively characterize the dynamic open conformation of the apoprotein. A brief summary of all three projects is given in chapter 7.

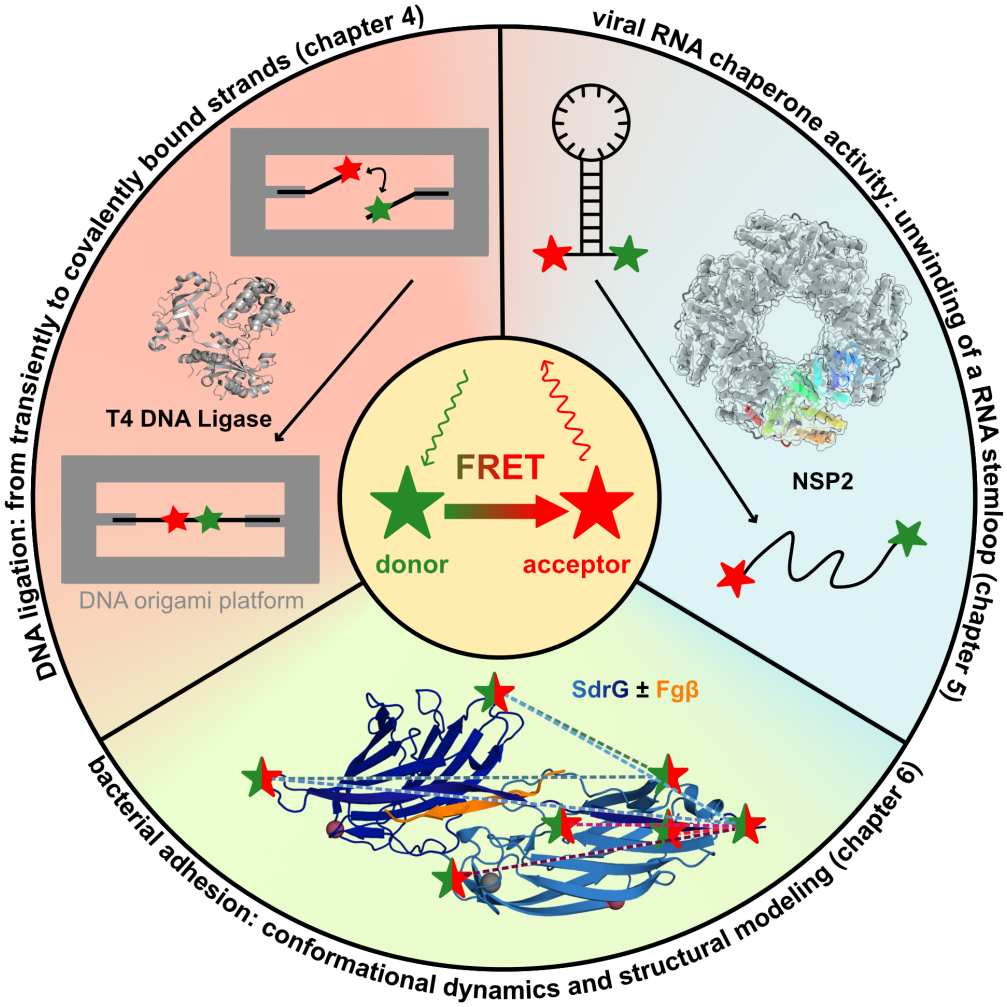


FIGURE 1.1: Overview of the proteins investigated in this thesis by smFRET. In the first project (chapter 4), the T4 DNA ligase (protein data base (PDB): 6DT1, Shi et al., 2018) is used to study the end-joining of two fluorescently labeled DNA double strands. To locally increase the concentration of the repair substrates, they are bound to a DNA origami platform. The unwinding of a dual-labeled RNA stemloop by the viral RNA chaperone NSP2 is studied in a second project (chapter 5). Shown is a cryoEM reconstruction of octameric NSP2 with a monomer highlighted in color, adapted from our NSP2-ΔC manuscript. In the third project (chapter 6) a network of FRET pairs is placed on bacterial adhesion protein SdrG (PDB: 1R17, Ponnuraj et al., 2003) to study conformational changes in the presence and absence of the target peptide Fgβ. Based on FRET-derived distances and MD simulations, a structural model of the dynamic apoprotein is proposed.

Chapter 2

Fluorescence

2.1 Basic principles of fluorescence

Photoluminescence is the emission of light after absorption of a photon. For light emitted from an electronically excited singlet state, the term fluorescence is used. The typical processes that occur between photon absorption and emission are illustrated in the Jablonski diagram in Figure 2.1. Here, the electronic states and their vibrational energy levels are depicted as horizontal lines. Upon absorption of a photon (represented by purple and blue arrows), a molecule is excited from its electronic ground state S_0 to a higher energetic singlet state S_n. This transition occurs on the femtosecond (10^{-15} s) timescale, which is too fast for the nuclei to significantly alter their positions. Thus, the transition probability from the vibrational ground state of S_0 to the different vibrational energy levels of S_n is mainly governed by the overlap of the vibronic wave functions (Franck-Condon principle, (Condon, 1926, 1928; Franck and Dymond, 1926)). Excitation is followed by internal conversion, which is the rapid relaxation to the lowest vibrational energy level of S_1. As all fluorescent transitions start from this energy level, fluorescence emission spectra do not depend on the excitation wavelength (Kasha's rule, (Kasha, 1950)). By returning to one of the vibrational energy levels of the ground state S_0, a photon (green arrow) is emitted. As the energy of the emitted photon is lower compared to the absorbed photon, fluorescence emission spectra are shifted to longer wavelengths (Stokes shift, (Stokes, 1852)).
Alternatively to fluorescence emission, molecules can relax non-radiatively to the ground state. Fluorescence quenching can be the result of a collision with the quencher (dynamic quenching) or the formation of a nonfluorescent complex (static quenching). Further, molecules in the excited state S_1 can undergo a spin conversion to the first triplet state T_1 by intersystem crossing. The slow return to the ground state S_0 (shown as orange arrows) is then termed phosphorescence.

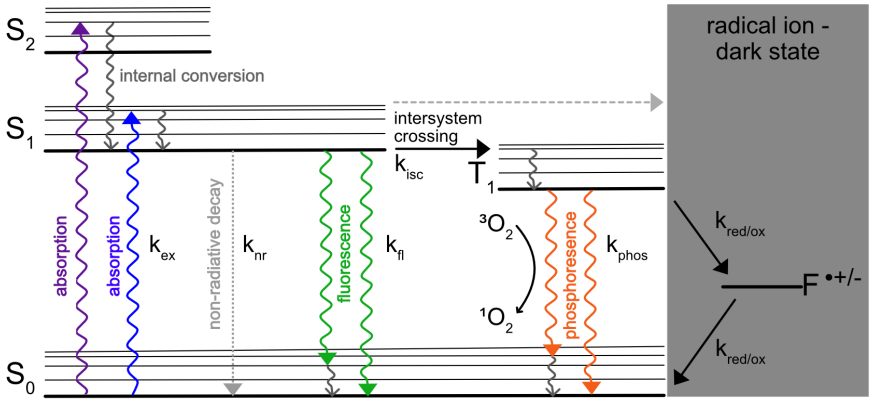


FIGURE 2.1: A Jablonski diagram illustrating the processes that occur during photoluminescence with S_0: singlet electronic ground state, S_1,S_2: singlet first and second electronic excited state, T_1: triplet state, k_{ex}: excitation rate, k_{nr}: non-radiative decay rate, k_{fl}: intrinsic fluorescence emission rate, k_{isc}: rate for intersystem crossing, and k_{phos}: phosphoresence emission rate. Horizontal lines represent the vibrational energy levels of each electronic state. The reaction with triplet oxygen (O_2^3) depopulates the triplet state and will be discussed further in section 2.2.2 of this chapter. The radical ion dark state $F^{\bullet+/-}$ is formed upon a reaction of the excited singlet or triplet states with an oxidizing or reducing agent and can be recovered by the complementary redox reaction.

2.1.1 Fluorescence lifetime

Important characteristics of a fluorophore are its fluorescence lifetime τ and quantum yield ϕ. τ describes the average time a molecule spends in the excited state before returning to the ground state. The intrinsic or natural lifetime τ_n is given by the inverse of the fluorescence emission rate k_{fl}

$$\tau_n = \frac{1}{k_{fl}} \tag{2.1}$$

As the return to the ground state can also occur through non-radiative processes, the measured lifetime τ is defined as

$$\tau = \frac{1}{(k_{fl} + k_{nr})} \tag{2.2}$$

where k_{nr} is the non-radiative decay rate. Typical fluoresence lifetimes are on the nanosecond (10^{-9} s) timescale. In contrast, phosphorescence lifetimes are much longer in the range of milliseconds (10^{-3} s) to seconds (Lakowicz, 2006), as the return from the triplet state T_1 to the singlet ground state S_0 is quantum mechanically forbidden. The quantum yield, ϕ, describes the fraction of photons that are emitted after an excitation event and depends on the radiative and non-radiative decay rates

$$\phi = \frac{k_{fl}}{(k_{fl} + k_{nr})} = \frac{\tau}{\tau_n} \tag{2.3}$$

thus defining the efficiency of the fluorescence process.

2.1.2 Fluorescence anisotropy

The fluorescence anisotropy is a measure of the changing orientation of a molecule in the time between absorption and emission events and contains information about the rotational freedom of fluorophores. When molecules are exposed to polarized light (e.g. a laser beam), the absorption probability is highest for photons whose polarization is aligned parallel to the absorption dipole moment of the molecule. Due to rotational diffusion, the orientation of the fluorophore can change during the lifetime of the excited state. The polarization of the emitted light in the plane defined by the dipole moment and the direction of propagation of the photon thus differs from the polarization of the excitation source and their correlation is measured by the fluorescence anisotropy r. Averaged over the emission time of the fluorophore, the steady-state anisotropy is defined by

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{2.4}$$

where $I_{\parallel}$ and $I_{\perp}$ are the fluoresence intensities of the parallel and polarized emission with respect to the polarization of the excitation light. The anisotropy of the emitted light depends on the distribution of the angle θ between the emission dipole moment of the molecule and its z-axis according to

$$r = \frac{3\langle\cos^2\theta\rangle - 1}{2} \tag{2.5}$$

with $\langle\rangle$ denoting the average. The excitation of molecules in different orientations results in $\cos^2\theta = \frac{3}{5}$ and a maximum anisotropy value of $r_{max} = 0.4$. Small molecules, like organic dyes, rotate fast in solution which leads to depolarized emission. As the orientational information is lost, anisotropy values tend to zero. When a dye is attached to a large biomolecule such as a protein via a rigid linker, its rotational diffusion is considerably slower and higher anisotropy values close to $r_{max} = 0.4$ are obtained.

2.2 Fluorophores

Typical fluorophores are aromatic molecules with conjugated π bonds, which occur naturally in some biomolecules. For example, the indole group of the amino acid tryptophan acts as an intrinsic fluorophore in proteins. For other molecules like deoxyribonucleic acid (DNA), which show no or only weak fluorescence, extrinsic fluorophores can be attached or a fluorescent nucleobase analog used. The experiments presented in this thesis are based on biomolecules that were labeled with extrinsic fluorophores such as the organic dye Atto647N.

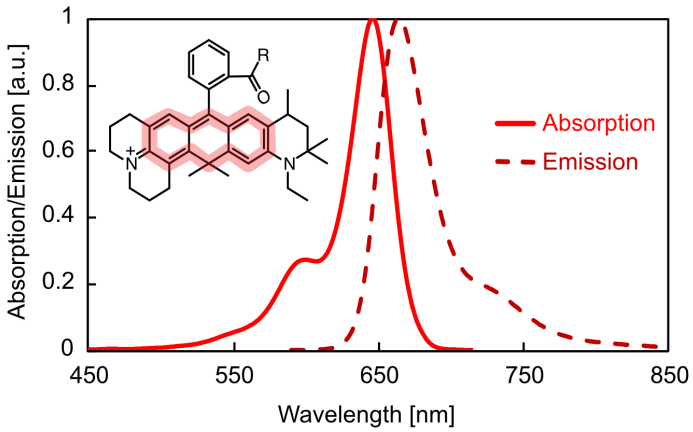


FIGURE 2.2: Absorption (solid line) and emission spectrum (dashed line) of the organic fluorophore Atto647N in phosphate buffered saline (PBS) based on data provided by the manufacturer at www.atto-tec.com (accessed April 2, 2020). The molecular structure of the dye contains a conjugated aromatic system, which is highlighted in red.

The absorption and emission spectra of this fluorophore are shown in Figure 2.2 with the fluorescent moiety of the molecular structure highlighted in red. The shape of the emission and absorption spectra are similar, but show an inverse energy dependence. This mirror image can be explained by similar vibrational structures of the ground state S_0 and the excited state S_1 with quick vibrational and rotational relaxation times. As relaxation from S_1 can occur to higher vibrational levels of S_0, the energy of the emitted photons is lower compared to the absorbed photons and the emission spectrum thus shifted to longer wavelengths (Stokes shift).

2.2.1 Conjugation of fluorophores to biomolecules

A variety of strategies exists to attach a fluorophore to a biomolecule of interest. The approaches relevant for this thesis are presented below.
Nucleic acids were fluorescently labeled by the manufacturer post-synthetically. The labels were introduced internally or at the 3′- or 5′-terminus by using amino-modified nucleotides. The underlying chemical reaction is identical for the different labeling positions: a fluorophore with a NHS ester moeity reacts with the amino group of the modified nucleotide to form an amide bond (Figure 2.3A). For the internal labeling of the DNA strands used in the DNA origami project (chapter 4), an internal thymidine was replace by the modified variant 5-C6-Amino-2′-deoxythymidine, which links the amino group to the nucleobase (Figure 2.3B). For terminal labeling of an RNA stemloop (chapter 5), amino-linkers with a C6-spacer arm were attached directly to the 3′-or 5′-end (Figure 2.3C and D).
Protein labeling was performed via thiol-maleimide chemistry (Figure 2.4). Here, the thiol group of the amino acid cysteine is targeted by a maleimide-functionalized dye. This reaction was used to stochastically label double-cysteine mutants of the bacterial adhesion protein SdrG with two different dyes (chapter 6).
The length of the linker and its attachment point determine the exact position of the fluorophore. This information is critical when the experimentally derived distances are interpreted with respect to the molecular structure and will be discussed further in section 2.3.1 of this chapter.

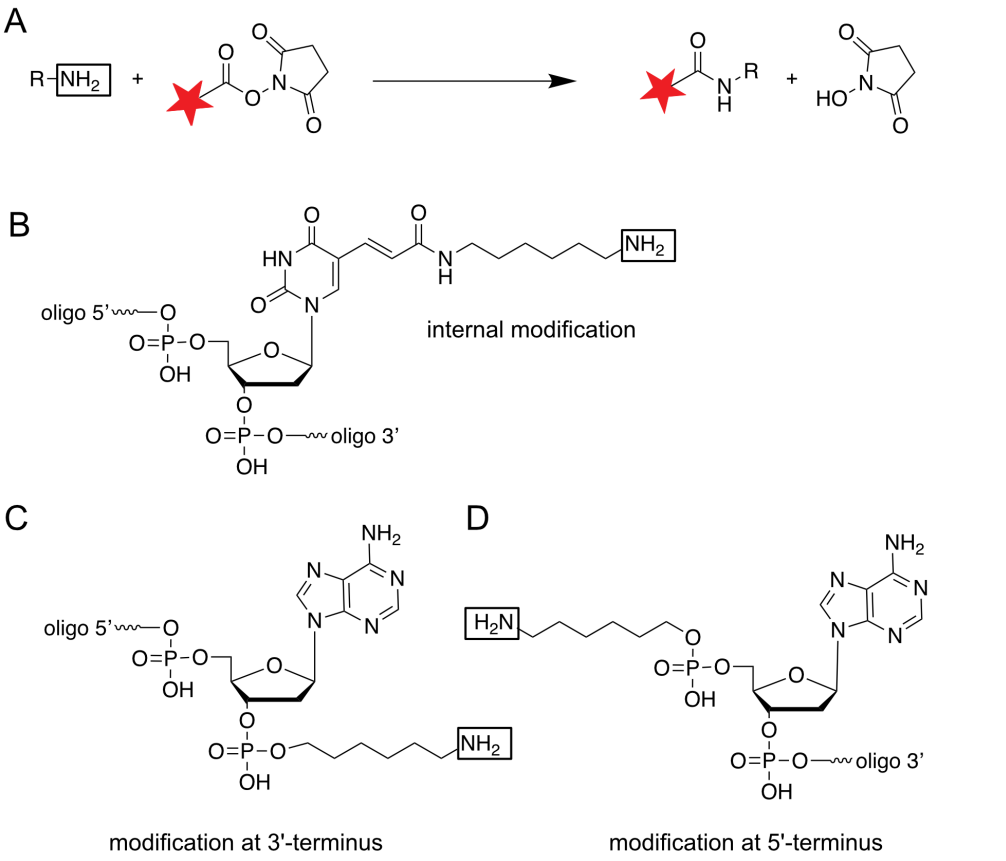


FIGURE 2.3: Reaction of an NHS ester with an amino group for DNA labeling. **A** A fluorophore (shown as a red star) with a NHS ester modification is conjugated to an amino-modified nucleotide. **B** For internal labeling, a nucleotide with an amino-linker attached to the nucleobase is used. Shown here is 5-C6-Amino-2′-deoxythymidine. **C** and **D** For terminal labeling, the amino-linker is directly attached to the 3′- or 5′-end of the oligonucleotide. Shown here is a C6-spacer, which is typically used to enable free rotation of the fluorophore on the biomolecule.

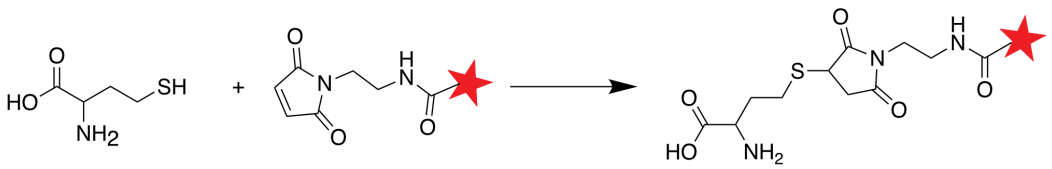


FIGURE 2.4: Thiol-maleimide reaction for protein labeling. The thiol group of the amino acid cysteine is fluorescenty labeled with a maleimide-functionalized dye.

2.2.2 Photostabilizing agents

Organic fluorophores often show blinking due to fluctuations between fluorescent and dark states like the triplet state, or fast photobleaching associated with the permanent destruction of the dye structure. To improve photostability, the triplet state has to be quenched and reactions with photodamaging agents avoided.
As triplet oxygen (O_2^3) can effectively depopulate the triplet state by forming the higher energetic singlet oxygen species (O_2^1) (Hübner et al., 2001), it reduces blinking. However, singlet oxygen damages the structure of organic dyes and biomolecules (Davies, 2004; Sies and Menck, 1992). To avoid photobleaching and maintain the stability and function of the labeled molecules, oxygen is typically removed from aqueous buffer solutions. The two most commonly used enzymatic oxygen scavenging systems are based on protocatechutic acid (PCA) and protocatechute-3,4-dioxygenase (PCD) (Aitken et al., 2008) or the combination of glucose oxidase and catalase (GODCAT) with a glucose-containing buffer (Stein et al., 2012). To quench the triplet state in the absence of oxygen, a combination of reducing and oxidizing agents is required to form a radical ion by electron transfer, which can subsequently be recovered through the complementary redox reaction (Vogelsang

et al., 2008) (Figure 2.1). The reducing and oxidizing system (ROXS) chosen for the experiments presented in this thesis is the vitamin E analogue 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) and its oxidized form, Trolox quinone. To generate the quinone species, a buffer solution containing 1 mM Trolox[1] was illuminated under a UV lamp (at wavelenghts of 254 nm and 366 nm). Over time, more and more Trolox quinone formed as indicated by the shift of the absorption maxima (Figure 2.5A). The percentage of Trolox quinone in solution was calculated using the absorbance at 255 nm as described in Cordes et al., 2009. After 4 min of UV illumination, the solution contained 10 % Trolox quinone, which lies well within the recommended range of 2-30 %, and the fraction increased only slightly with additional illumination time (Figure 2.5B). An illumination time of 5 min was thus chosen to be sufficient to prepare Trolox/Trolox quinone buffers.

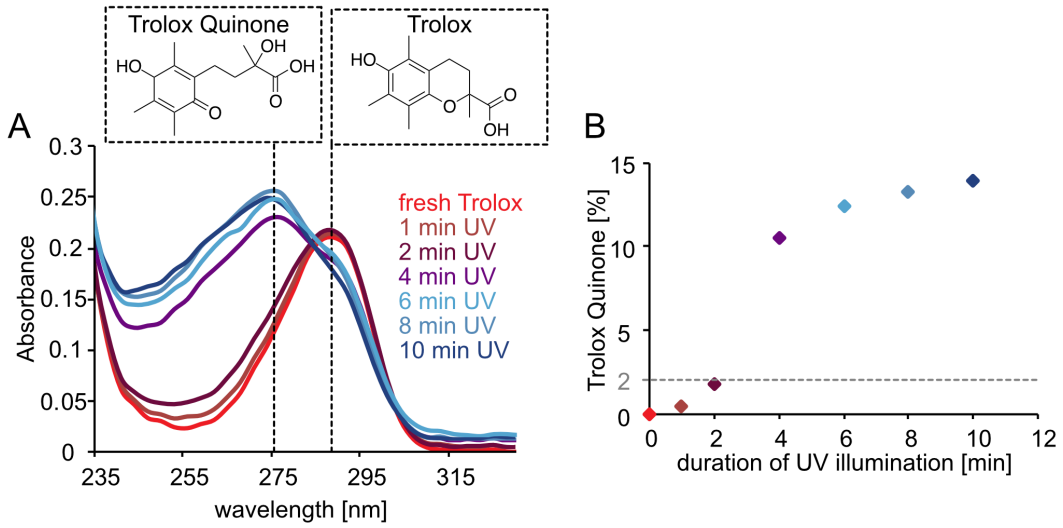


FIGURE 2.5: UV light-induced oxidation of Trolox to Trolox quinone generates a photostabilizing ROXS solution within minutes. **A** Absorption spectra of 1 mM Trolox in water before and after illumination with UV light (at 254 nm and 366 nm) through an eppendorf tube. Trolox has an absorption maximum at $\sim$ 290 nm and the oxidized quinone appears at $\sim$ 275 nm. The longer the illumnation time, the more Trolox is oxidized. **B** The fraction of Trolox quinone (calculated via the absorbance at 255 nm as described in Cordes et al., 2009) increases with illumination time. An illumination time of 4 min or longer generates enough Trolox quinone for an efficient ROXS buffer (recommended: 2-30 % Trolox quinone).

2.3 Förster resonance energy transfer

When two fluorophores are sufficiently close together, energy can be transferred from a donor molecule in the excited state to an acceptor molecule in the ground state by dipolar coupling, as depicted in Figure 2.6A. This radiationless transfer, known as Förster Resonance Energy Transfer (FRET, (Förster, 1948)), requires an overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor and is based on dipole-dipole interactions. By placing both FRET dyes on a biomolecule, conformational changes can be probed (intra-molecular FRET). When the FRET dyes are positioned on two different molecules, their dynamic interaction can be detected (inter-molecular FRET). The rate of the energy transfer k_T strongly depends on the distance R between the two fluorophores and is given by

$$k_T = \frac{1}{\tau_{D(0)}} \left(\frac{R_0}{R}\right)^6 \tag{2.6}$$

with the fluorescence lifetime of the donor in absence of the acceptor, $\tau_{D(0)}$, and the Förster radius R_0. The efficiency of the energy transfer is calculated as the ratio of the transfer rate to the total decay rate of the donor in presence of the acceptor:

$$E = \frac{k_T}{k_T + \tau_D^{-1}} \tag{2.7}$$

[1] freshly diluted from a 100 mM stock in ethanol

which can be rewritten as

$$E = \frac{R_0^6}{R_0^6 + R^6} = \frac{1}{1 + (\frac{R}{R_0})^6} \tag{2.8}$$

The Förster radius R_0 is the inter-dye distance that results in an energy transfer efficiency of 50 %, and is given by

$$R_0^6 = \frac{9000(\ln 10)\kappa^2 \phi_D}{128\pi^5 N_A n^4} J(\lambda) \tag{2.9}$$

where κ^2 is a geometric factor that depends on the relative orientation of the acceptor and donor transition dipoles, ϕ_D is the quantum yield of the donor, N_A is Avogadro's Number, n is the refractive index of the medium, and $J(\lambda)$ is the overlap integral of the acceptor's absorption and the donor's emission spectrum. $J(\lambda)$ depends on the normalized emission spectrum of the donor $\int_0^\infty F_D(\lambda)d\lambda = 1$ and the absorption spectrum of the acceptor expressed in terms of the extinction coefficient $\epsilon_A(\lambda)$:

$$J(\lambda) = \int_0^\infty F_D(\lambda)\epsilon_A(\lambda)\lambda^4 d\lambda \tag{2.10}$$

The orientation factor κ^2 is calculated as

$$\kappa^2 = (\cos\theta_T - 3\cos\theta_D \cos\theta_A)^2 \tag{2.11}$$

with the angle between the donor emission dipole moment and the acceptor absorption dipole moment, θ_T. θ_D and θ_A are the angles between the respective dipole and the vector joining the donor and acceptor. κ^2 can adopt values between 0 (for a perpendicular orientation of the dipoles) and 4 (for a head-to-tail parallel orientation of the dipoles). For freely rotating dyes with rotation times $< 1ns$ and lifetimes $\sim 4ns$, κ^2 is usually averaged over all possible angles to a value of $\frac{2}{3}$.
The distance-dependence of the FRET efficiency is exemplarly shown in Figure 2.6B for the dye pairs Alexa488-Atto647N, Atto 532-Atto647N and Cy3b-Atto647N with Förster radii of 54 Å, 59 Å and 67 Å, respectively. For larger values of R_0, larger distance separation can be sensitively probed.

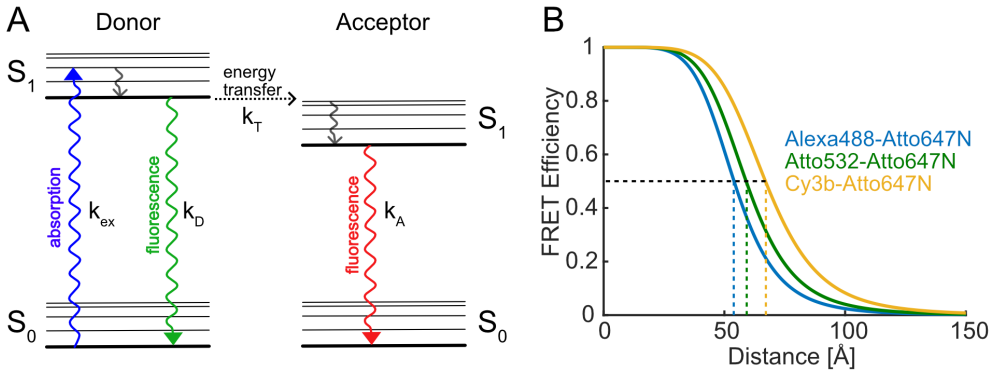


FIGURE 2.6: **A** A Jablonski diagram illustrating FRET between a donor and an acceptor dye with S_0: singlet ground state, S_1: first excited singlet state, k_{ex}: excitation rate, k_D and k_A: emission rates of the donor and acceptor fluorophores, respectively, and k_T the rate of the energy transfer. **B** Distance-dependent FRET efficiencies are shown for three dye pairs with different Förster radii, which are indicated by the colored dashed lines.

2.3.1 Finding optimal label positions for FRET

Labelling positions for FRET dyes on a biomolecule have to be chosen carefully. On the one hand, the introduction of the labels should not perturb the structure and functionality of the molecule. Thus, placing dyes in catalytically active or structurally crucial parts (e.g. structural motifs required for binding of interaction partners) should be avoided (Sánchez-Rico et al., 2017). On the other hand, the inter-dye distance has to lie within the FRET-sensitive range to detect conformational changes of the biomolecule. The inter-dye distance can be estimated, if the structure of the molecule is known (e.g. if its structure is deposited in the protein data bank (PDB)) or can be predicted (from structural modeling e.g. using the Nucleic Acid Builder program (Ivani

et al., 2016), from homology modeling, or from computer simulations). By modelling the dye as a sphere, sterically allowed positions that do not collide with the surface of the biomolecule can be calculated for individual attachment points. Importantly, the length of the linker between attachment point and dye molecule (usually around 20 Å when fully extended) and its width have to be included in the model. When all allowed positions are considered to be equally probable, an accessible volume (AV) can be defined for the dye (Sindbert et al., 2011). To calculate the fluorophores' AVs for the molecules presented in this thesis, the FRET positioning and screening (FPS) software developed in the Seidel lab (Kalinin et al., 2012) was used. For the SdrG protein presented in chapter 6, the C_α-atom of the amino acid cysteine was defined as the attachment point, for the internally labeled DNA strands in chapter 4, it was the C7 atom of the methyl group of the modified thymine base. The specific dye and linker parameters are described in the respective materials and method sections. Using the FPS software and a known Förster Radius R_0, FRET-averaged distances between the dyes, $\langle R_{DA} \rangle_E$, and mean FRET efficiencies $\langle E \rangle$ were calculated according to

$$\langle R_{DA} \rangle_E = R_0 (\langle E \rangle^{-1} - 1)^{\frac{1}{6}} \tag{2.12}$$

and

$$\langle E \rangle = \left\langle \frac{1}{1 + \frac{R_{DA}^6}{R_0^6}} \right\rangle \tag{2.13}$$

where R_{DA} is the distance between donor and acceptor positions $|R_D - R_A|$. The theoretical distance and FRET efficiency values allow one to evaluate whether the inter-dye separation is in the FRET-sensitive regime and are thus a useful tool to identify appropriate label positions for FRET experiments. In addition, a comparison of experimental and theoretical distances (e.g. extracted from a crystallographically resolved protein structure) can be used to verify the experimental findings or to propose novel structures using the FRET information.

2.3.2 Ensemble versus single-molecule FRET

FRET experiments, which are performed at high molecular concentrations, usually result in high signal intensities above the background and do not require particularly sensitive equipment. However, as only an average FRET efficiency value is obtained, it is not possible to distinguish potential subspecies in bulk measurements. These heterogeneities can arise from biomolecules undergoing conformational changes or simply from an incomplete labeling reaction. To avoid ensemble averaging, molecules can be detected individually. By measuring only one molecule at a time, single-molecule FRET (smFRET) experiments allow not only the identification of subpopulations or rare events, but also the recording of time courses to determine kinetic rates of individual molecules. The first demonstration of FRET between a single donor and acceptor pair was reported almost 25 years ago (Ha et al., 1996). Numerous smFRET studies of various biological systems have been published ever since (Lerner et al., 2018) and it has been recently shown that intra- and intermolecular distances can be measured with sub-nanometer precision (Hellenkamp et al., 2018). Throughout this thesis, the term smFRET will be used for donor and acceptor dyes placed on the same biomolecule, while single-pair FRET (spFRET) will denote labels on two interaction partners.

To detect single molecules, sensitive photon detectors have to be used. Further, the background noise has to be sufficiently suppressed by reducing the observation volume. The two main microscope techniques that allow the realization of a small detection volume and that were applied for the smFRET experiments presented in this thesis are confocal and total internal reflection fluorescence microscopy. Both techniques are introduced in the following chapter together with a detailed description of the experimental procedures and the data analysis.

Chapter 3

Observation of single molecules

High sensitivity and low background are prerequisites for the detection of single-molecule events. To minimize the background signal and limit the necessary sample concentration, a small observation volume is required, which can be either realized on a confocal or a total internal reflection fluorescence (TIRF) microscope. On a confocal microscope, a laser is focused to a diffraction-limited spot, creating an observation volume of approximately 1 fL (10^{-15} L). Molecules diffusing into this confocal spot are excited, and any fluorescence signal from outside the desired volume is discarded by the pinhole in the detection path. In TIRF microscopy, an evanescent field with a limited penetration depth on the order of 100 nm exclusively excites molecules located close to the surface. This approach enables longer observation times, but requires the immobilization of the molecules to the surface. Both techniques, their experimental realization and data analysis are described in the following chapter.

3.1 Single molecules in solution

Single molecules can be measured in solution by keeping the observation volume of a confocal microscope fixed to a specific spot. Every time a fluorescently-labeled molecule diffuses into the confocal volume, a burst of photons is emitted. For sample concentrations on the order of 1-100 nM ($10^{-9} - 10^{-7}$ M), a correlation analysis of the time-dependent fluoresence intensity fluctuations reveals parameters like average diffusion times or the number of particels in the confocal volume. Decreasing the sample concentration further to the pM regime (10^{-12} M) allows a burst-wise analysis of individual molecules. In this thesis, fluorescence correlation spectroscopy (FCS) and burst-wise smFRET experiments were performed on confocal microscopes with pulsed-interleaved excitation (PIE, (Müller et al., 2005)) and multiparameter fluorescence detection (MFD, (Widengren et al., 2006)), which are described in detail below.

3.1.1 Confocal microscopy

The prototype of a confocal microscope was patented almost 60 years ago by Marvin Minsky (Minsky, 1961) and is schematically presented in Figure 3.1A. For excitation, a light source is focused by an objective lens into the sample. As Minsky did not use the collimated beam of a laser, a pinhole was required to limit the excitation to a small region in the focal plane. The fluorescence signal is collected by the same lens and separated from the excitation wavelength with a dichroic mirror. By placing a second pinhole in the detection path in a focal plane identical to that of the excitation pinhole (thus con-focal), any signal coming from below or above the focus is blocked and not detected as indicated by the dashed lines in Figure 3.1A. The lateral resolution in a confocal microscope is limited by the diffraction of light when focused by the objective lens. The diffraction image of a point emitter is described by the Airy disk and the lateral resolution limited by the Rayleigh criterion (Rayleigh, 1879)

$$r_{\text{lateral}} = \frac{0.61\lambda}{NA} \tag{3.1}$$

with the excitation wavelength λ and the numerical aperture of the objective lens, NA, which is defined as

$$NA = n \sin \alpha \tag{3.2}$$

n is the refractive index of the immersion medium (water, oil, or air) and α the maximum half-opening angle of the objective. The lateral resolution is approximated by the minimal distance at which two point emitters can still be distinguished, which requires an overlap of the minimum of the first emitter's Airy disk with the

maximum of the second emitter's Airy disk. The confocal volume, also termed point spread function (PSF), is the product of the excitation and detection Airy disks. For one-photon excitation, the PSF can be approximated by a three-dimensional Gaussian function according to

$$PSF = \exp\left[-2\left(\frac{x^2+y^2}{\omega_r^2} + \frac{z^2}{\omega_z^2}\right)\right] \tag{3.3}$$

where ω_r and ω_z are the distance from the PSF in the lateral (x, y) and axial (z) direction, respectively, at which the intensity has decayed to $1/e^2$ of the maximum value.
Although originally developed for imaging immobilized structures by scanning the sample, the observation volume can also be kept stationary to measure freely diffusing molecules in a dilute solution. The two confocal microscope setups that were used for diffusion-based experiments presented in this thesis are described in the following section.

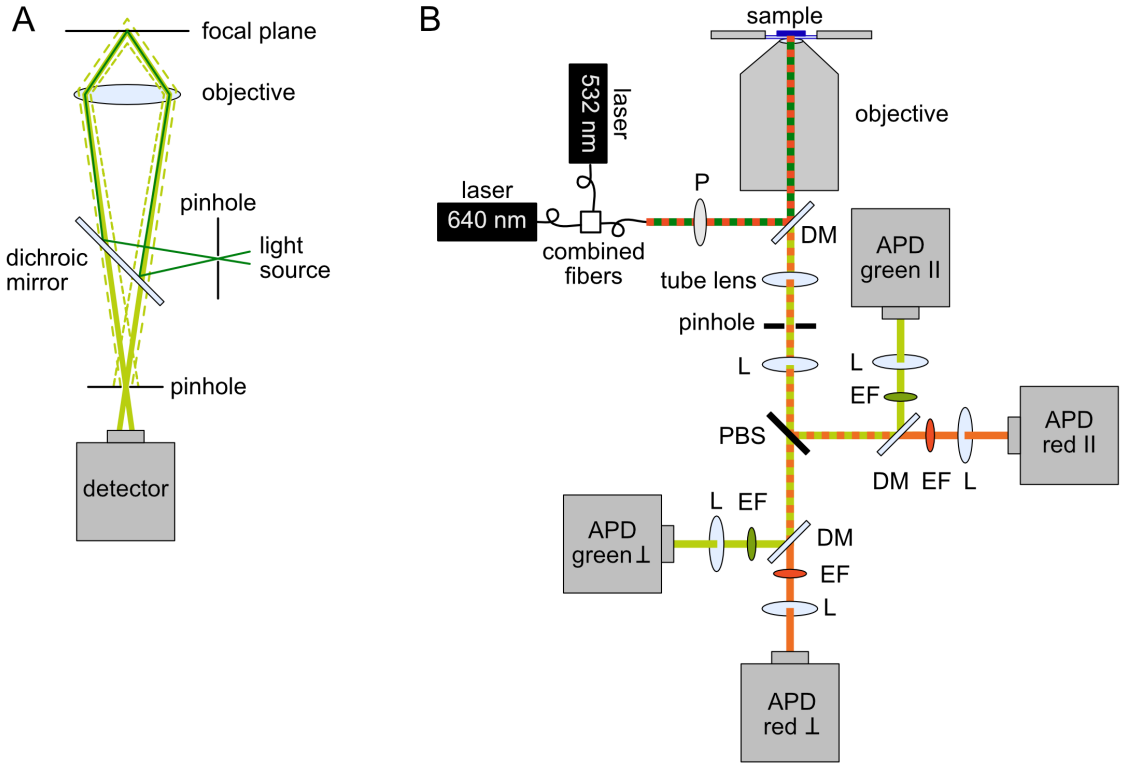


FIGURE 3.1: Confocal microscopy. **A** A prototype of a confocal microscope. The light source (dark green) is focused into the sample by the first pinhole. Only fluorescence signal originating from the focal plane (light green, solid line) passes through the second pinhole in front of the detector, while out-of-focus light (light green, dashed lines) is blocked. **B** A schematic representation of the two-color PIE-MFD confocal setup used in this work with an infinity-corrected objective and a tube lens to focus the fluorescence signal onto the pinhole. P: polarizer, DM: dichroic mirror, L: lens, PBS: polarizing beam splitter, EF: emission filter, and APD: avalanche photodiode for parallel (∥) and perpendicular (⊥) detection.

3.1.2 Experimental setups

Experiments were perfomed on two PIE-MFD confocal setups: Molecules labeled with Atto532 as FRET donor (i.e. the RNA stem-loop in chapter 5 and the SdrG protein in chapter 6) were measured on the two-color PIE-MFD microscope, while molecules labeled with Alexa488 or Cy3b as FRET donor (i.e. the DNA origami platform in chapter 4, the DNA-RNA hybrid structures in chapter 5 and the SdrG protein in chapter 6) were measured on the three-color PIE-MFD microscope.

Two-color PIE-MFD microscope

A schematic of the two-color PIE-MFD setup, which was originally developed a few years back in our research group (Kudryavtsev et al., 2012), is shown in Figure 3.1B.

For pulsed-interleaved excitation (PIE, (Müller et al., 2005)), laser lines at 532 nm (PicoTA 530, Toptica, Munich, Germany) and 640 nm (LDH-D-C-640, Picoquant, Berlin, Germany) are coupled into single-mode fibers (Schäfter+Kirchhoff, Hamburg, Germany), combined through a T-geometry wavelength division multiplexer (WDM-12P-111-532/647-3.57125-PPP-50-3A3A3A-3-1,1,2, OZ Optics, Carp, Canada) and collimated (60FC-4-RGB11-47,Schäfter+Kirchhoff). The pulsed lasers are operated at a repetition rate of $f = 26.67$ MHz (leading to pulses of 37.5 ns) with a delay of $\Delta t = 18$ ns. The laser light is cleaned by a Glan-Thompson polarizer (GTHM polarizer, Thorlabs, Dachau, Germany) and focused into the sample by a 60x water immersion objective (Plan Apo IR 60x/1.27 WI, Nikon, Düsseldorf, Germany). Laser powers before the objective of 100 μW were typically used.
For the simultaneous detection of various fluorescence parameters like intensity, lifetime and anisotropy (multiparameter fluoresence detection (MFD) (Widengren et al., 2006)), fluorescence is collected by the same objective, seperated by a dichroic mirror (DualLine z532/635, AHF Analysentechnik, Tübingen, Germany), focused through a pinhole with a diameter of 75 µm, and parallel and perpendicular components seperated by a polarizing beam splitter (05FC16PB.3, Newport, Darmstadt, Germany). For both parallel and perpendicular signals, fluorescence is spectrally split by a dichroic mirror (640DCXR, AHF Analysentechnik) and cleaned by emission filters (Brightline HQ582/75 for the green signal and Brightline HQ700/75 for the red signal, both from AHF Analysentechnik). Signals are detected on four single-photon counting avalanche photodiodes (APDs, SPCM-AQR, Perkin Elmer, Hamburg, Germany), each of which is connected to a seperate time-correlated single photon counting (TCSPC) card (SPC-154, Becker & Hickel GmbH, Berlin, Germany).

Three-color PIE-MFD microscope

The three-color PIE-MFD microscope is similar to the two-color setup shown in Figure 3.1B, but allows the use of up to three laser excitation lines at 482 nm, 565 nm and 641 nm (LDH-D-C-485, LDH-D-TA-560, and LDH-D-C-640, Picoquant), which are operated at a repetition rate of $f = 16.67$ MHz (60 ns pulses) with a delay of $\Delta t = 20$ ns. The fluorescence signal is collected by a 60x water immersion objective (Plan Apo IR 60x/1.27 WI, Nikon, Düsseldorf, Germany) and focused through a pinhole with a diameter of 50 µm. Typical laser powers before the objective of 100 μW for blue, 80 μW for green and 40 μW for red excitation were used. Parallel and perpendicular signals are seperated by a polarizing beam splitter, spectrally split by a set of two dichroic mirrors (BS560, 640DCXR, AHF, Analysentechnik), spectrally selected by emission filters (ET525/50 for blue and ET607/36 for green, AHF, Analysentechnik) and focused on six single-photon counting APDs. Photon arrival times were recorded at 16 ps resolution using TCSPC electronics (HydraHarp400, Picoquant, Berlin, Germany). As only two-color experiments with either red and green or red and blue excitation were performed during this thesis, different red emission filters were used, depending on the second wavelength. For blue excitation at 482 nm, ET705/100 emission filters (AHF, Analysentechnik) were used in front of the red detectors. For green excitation at 565 nm, ET667/30 and ET670/30 filters (AHF, Analysentechnik) were used for the parallel and perpendicular red detectors, respectively, to avoid Raman background.

3.1.3 Pulsed-interleaved excitation (PIE)

To detect a FRET signal of a freely diffusing molecule in solution, it is in principle sufficient to excite the donor fluorophore and spectrally seperate the emission into the donor and acceptor channels. However, the absence of an acceptor signal after donor excitation can have several reasons: Either the distance between donor and acceptor dyes is indeed too large for an efficient energy transfer, the acceptor fluorophore is not present due to an incomplete labeling reaction or it is permanently or temporary non-fluorescent due to photobleaching or blinking. Pulsed-interleaved excitation (PIE, (Müller et al., 2005)), schematically presented in Figure 3.2, directly probes the presence and photoactivity of the acceptor fluorophore. In this method, two or more lasers are operated at identical frequencies f and their nanosecond pulses delayed by a short time window

$$\Delta t = \frac{1}{nf} \tag{3.4}$$

where n is the number of excitation lasers used. Each excitation cycle is electronically triggered by a synchronization pulse (sync pulse). Using TCSPC detection, photons are not only sorted into different detection channels, but also according to their excitation source. For experiments with two excitation lasers, four PIE channels are thus definded: green signal after green excitation (GG), red signal after red excitation (RR), red signal after green excitation (GR) and green signal after red excitation (RG). The latter PIE channel is typically ignored, as the signal after red excitation is efficiently blocked by the green emission filters.

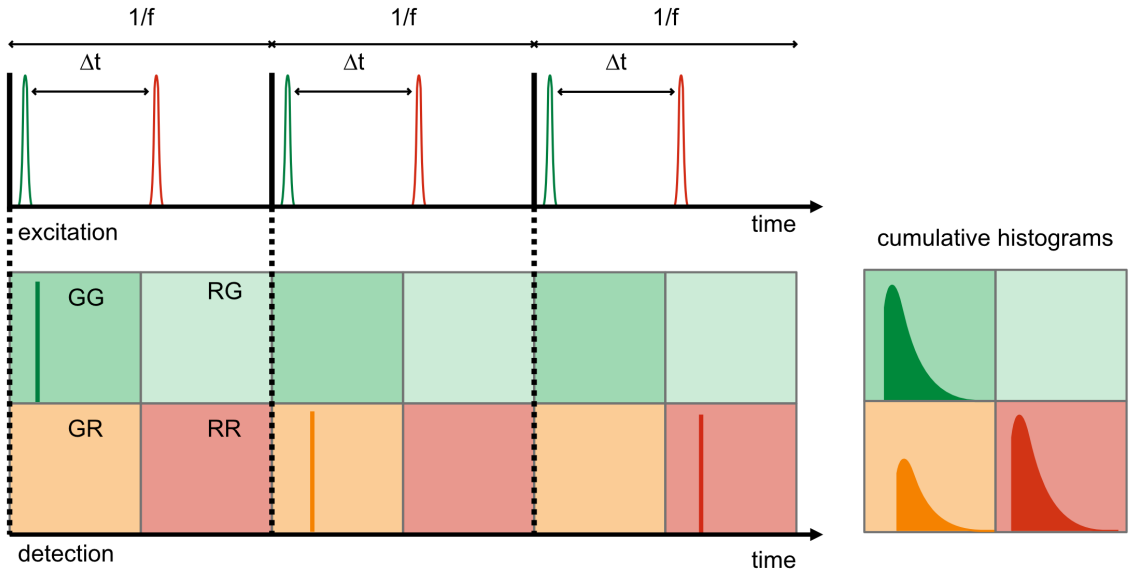


FIGURE 3.2: Excitation and detection in PIE experiments. Upper part: For excitation, two lasers are operated at an identical frequency f and their nanosecond pulses delayed with respect to each other by a short time period Δt. Lower part: Photons are detected in the four PIE channels *GG*, *RG*, *GR*, and *RR*. The first letter indicates the excitation color and the second letter the detection channel. All recorded photons contribute to the fluorescence lifetime decays shown in the cumulative histograms.

The channel *GR* does not only contain the FRET signal, but also contributions from spectral crosstalk of the green dye (α) as well as from direct excitation of the red dye by the green laser (δ). Using PIE, all neccessary correction factors can be directly determined from the data set, as described in Section 3.1.6.

3.1.4 Fluorescence correlation spectroscopy (FCS)

For this thesis, fluorescence correlation spectroscopy (FCS) experiments were mainly performed to characterize the diffusion behavior of fluorescently labeled molecules.
The method is based on time-dependent fluorescence intensity fluctuations of molecules diffusing in and out of the confocal volume and was first described in the early 1970s (Magde et al., 1972). The temporal signal fluctuations are analyzed using the autocorrelation function (ACF), which assesses the self-similarity of the signal as a function of the lag time τ. The ACF is given by

$$G(\tau) = \frac{\langle \delta I(t) \delta I(t+\tau) \rangle}{\langle I(t) \rangle^2} = \frac{\langle I(t) I(t+\tau) \rangle}{\langle I(t) \rangle^2} - 1 \tag{3.5}$$

where $I(t)$ is the fluorescence signal at time t, $\delta I(t)$ the deviation from the mean intensity, and $\langle\ \rangle$ the time average. The ACF amplitude depends on the number of molecules in the focus, while its time-dependent decay reports on the time scale of molecular motion. Assuming a Gaussian-shaped focal volume (see Equation 3.3), the ACF for freely diffusing particles in 3D is definded as

$$G(\tau) = \frac{\gamma}{\langle N \rangle} \frac{1}{1 + \frac{\tau}{\tau_D}} \frac{1}{\sqrt{1 + \frac{1}{p^2} \frac{\tau}{\tau_D}}} \tag{3.6}$$

where $\gamma = 2^{-3/2}$ accounts for the non-uniform illumination in the 3D Gaussian volume. $\langle N \rangle$ is the average particle number, and τ_D the diffusion time, which is connected to the diffusion coefficient D by

$$\tau_D = \frac{\omega_r^2}{4D} \tag{3.7}$$

with the lateral dimension of the confocal volume ω_r. Finally, the p-factor accounts for different sizes of the confocal volume in the lateral and axial dimensions according to

$$p = \frac{\omega_z}{\omega_r} \tag{3.8}$$

To determine the diffusion coefficient D, the size of the confocal volume needs to be known. To this end, a solution of fluorophores with known diffusion coefficients (Kapusta, 2010) was measured on the beginning of each day and the confocal volume parameters were extracted from the autocorrelation functions. The molecules of interest were then typically diluted to 1 nM and measured in eight-well chamber slides (Nunc Lab-Tek 1.0 borosilicate glass, VWR), which were coated with bovine serum albumin (BSA, 1 mg/ml) to reduce sticking to the glass surface. Dye and sample solutions were usually measured for 2 minutes and 5-10 minutes, respectively. Data were analyzed using the open-source software *PAM* (PIE Analysis in Matlab) (Schrimpf et al., 2018) that was developed in our group. Typical autocorrelation curves for the free dye Atto655 and for Atto647N-labeled DNA structures are shown in Figure 3.3. As the hydrodynamic radius and thus the diffusion time increases from the free dye to the DNA origami structure, the ACFs are shifted to longer time scales.

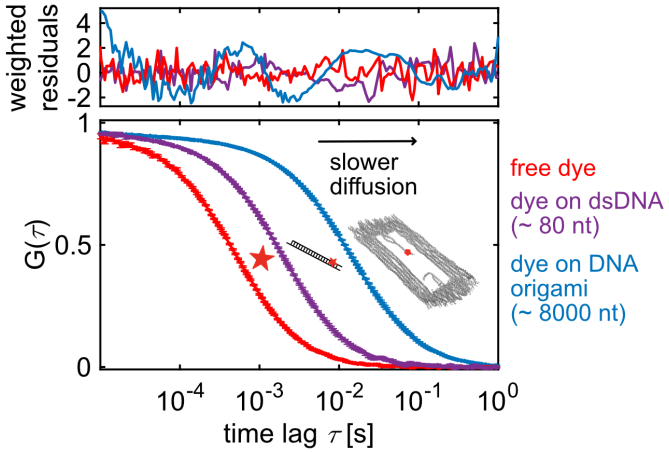


FIGURE 3.3: FCS reports on the diffusion time of fluorescently labeled molecules. Shown are amplitude-normalized ACFs of the dye Atto655 (in red, D = 373 $\mu m^2/s$), a DNA double strand labeled with Atto647N (in purple, D = 4 $\mu m^2/s$), and an Atto647N-labeled DNA origami structure (in blue, D = 12 $\mu m^2/s$) diffusing in solution and the weighted residuals of the fit. A larger molecular weight results in a slower diffusion and thus in a shift of the autocorrelation curve to longer timescales.

Intensity fluctuations in FCS experiments are not solely caused by molecular diffusion. Factors like photophysical effects, conformational dynamics or photon antibunching can further influence the autocorrelation function on shorter timecales (Felekyan et al., 2005; Felekyan et al., 2013; Nettels et al., 2007).

Fluorescence cross-correlation spectroscopy (FCCS)

A co-diffusion and thus interaction of two fluorescently-labeled binding partners can be revealed by a cross-correlation analysis (Schwille et al., 1997). The molecules of interest are labeled with different fluorophores and their signals spectrally separated into the detection channels i and j. The cross-correlation function (CCF) is given by

$$G_{CCF}(\tau) = \frac{\langle \delta I_i(t) \delta I_j(t+\tau) \rangle}{\langle I_i(t) \rangle \langle I_j(t) \rangle} \tag{3.9}$$

The CCF amplitude at $\tau = 0$ is proportional to the number of two-color molecules N_{ij} and inversely proportional to the total number of molecules with the first label, N_i, and the second label, N_j:

$$G_{CCF}(0) = \gamma \frac{N_{ij}}{N_i N_j} \tag{3.10}$$

A CCF amplitude larger than zero thus indicates co-diffusion of the dyes. However, even in the absence of molecular interactions, a non-zero CCF might be caused by spectral crosstalk from the shorter-wavelength fluorophore into the longer-wavelength detection channel. Using PIE, this crossstalk contribution can be removed and the corrected CCF used for a quantitative analysis of molecular interactions.

3.1.5 Burst-wise smFRET experiments using PIE-MFD

Having characterized the samples with FCS, we could next perform smFRET experiments. As mentioned above, FCS experiments are conducted at nM concentrations. By using even lower concentrations in the pM regime, fluorescence bursts originating from individual molecules can be sufficiently distinguished and a burst-wise analysis performed. Using PIE, double-labeled molecules can be separated from single-labeled species and correction factors determined to quantify the FRET efficiency of the single-molecule events. As the confocal setups used in this work are equipped with MFD, information on fluorescence lifetime and anistotropy are additionally available, which can give insights into conformational dynamics up to the time scale of diffusion, and the rotational diffusion of the molecules, respectively.
Two-color smFRET experiments on either the two-color or three-color PIE-MFD microscope were performed in eight-well chamber slides (Nunc Lab-Tek 1.0 borosilicate glass, VWR). To obtain the instrument response function (IRF), which characterizes the timing response of the system, a clean (i.e. non-fluorescent) PBS solution (purchased from ThermoFisher Scientific) was measured for 10 min. The purchased PBS solution was chosen to determine the IRF, as both freshly prepared buffer solutions or water from our MilliQ machine usually showed some minor fluorescent background due to contamination. To determine the fluorescent background, a buffer measurement was performed for 5 min. Fluorescently-labeled proteins or nucleic acids were then diluted to pM concentrations and recorded for 1-5 hours. To reduce sticking of the molecules to the glass surface, the Lab-Tek slide was either coated with BSA, or 0.01 % (v/v) Tween 20 added to the buffer. The measurement buffer usually contained 1 mM Trolox for photostabilization (Cordes et al., 2009) as mentioned in section 2.2.2 of the previous chapter. As oxygen was not removed from the buffer, a sufficient amount of Trolox was oxidized to its quinone form during the experiment and thus no activation by UV light required prior to the measurement.
SmFRET data were analyzed with the program *PAM* (Schrimpf et al., 2018) as decribed in the following section.

3.1.6 smFRET data analysis

Burst selection

Fluorescent bursts originating from individual molecules were selected using the all-photon search algorithm, which is based on a sliding time window approach (Nir et al., 2006). In this algorithm, the local count rate is determined for every detected photon to evaluate whether this photon is part of a burst. To define a burst, thresholds are set for the length of the sliding time window around each photon (tw), the minimum number of photons per time window (p_{tw}) and the minimum number of total photons per burst (p_{total}). Figure 3.4A illustrates two examples of the sliding time window search. In the upper panel, fluorescent bursts are shown for a large DNA origami structure, while the lower panel shows bursts of a small protein. Every burst above the chosen threshold (indicated by a dashed line) is selected for further analysis. Bursts of the DNA origami sample are simultaneously detected in the red and green channels. For the protein measurement, however, also one-color bursts are visible, which indicate donor- and acceptor-only species and thus an incomplete labeling reaction. It is possible to excluded these one-color species right from the beginning by performing a dual channel burst search (Nir et al., 2006). Nonetheless, the all-photon search was chosen for the data presented in this thesis, as the single-color bursts provide valuable information on the correction factors (Kudryavtsev et al., 2012) and can be removed at a later stage in the analysis.
The bursts of the DNA origami are considerably longer than the protein bursts. Thus, the sliding time window should be adjusted based on the diffusion time. Typical burst search parameters for DNA origami samples were $tw = 2500\,\mu s$, $p_{tw} = 50$ and $p_{total} = 500$. For proteins and other small molecules like DNA double strands or RNA stem-loops, parameters of $tw = 500\,\mu s$, $p_{tw} = 10$ and $p_{total} = 100$ were chosen.
The average burst duration of the DNA origami is around 10-fold slower compared to the protein sample (Figure 3.4B), while the molecular weights differ by a factor of 1000. This observation is in agreement with the Stokes-Einstein relation, which relates the hydrodynamic radius of a molecule r to its diffusion coefficient D:

$$D = \frac{k_B T}{6\pi\eta r} \tag{3.11}$$

with the Boltzmann constant k_B, the temperature T and the viscosity η. Assuming a spherical particel with volume V and mass m ($V = \rho m = \frac{4}{3}\pi r^3$), D is inversely proportional to the third root of the molecular mass

$$D \propto \frac{1}{r} \propto \frac{1}{\sqrt[3]{V}} \propto \frac{1}{\sqrt[3]{m}} \tag{3.12}$$

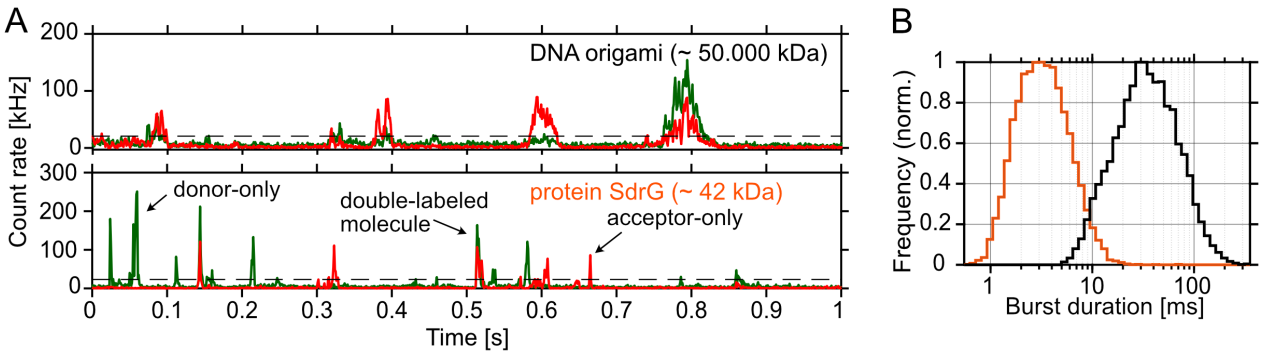


FIGURE 3.4: Parameters for the sliding time window burst search depend on the molecular diffusion time. **A** Fluorescent bursts of a double-labeled DNA origami structure (upper panel) and the protein SdrG (lower panel) detected in the red and green channels. The dashed lines indicate the threshold for burst selection. **B** Histograms of fluorescence burst duration with an average value of 3 ms for the protein SdrG (in orange) and 30 ms for the DNA origami (in black).

Thus, the diffusion time, which is inversely proportional to D (see Equation 3.7), scales with the third root of the molecular weight.

Correction factors for quantitative smFRET

Having identified bursts originating from single-molecule events, multiple parameters can be calculated. The uncorrected FRET efficiency or proximity ratio E_{PR} is directly accessible through the background-corrected intensities in the donor channel after donor excitation (I_{GG}) and in the acceptor channel after donor excitation (I_{GR}):

$$E_{PR} = \frac{I_{DA}}{I_{DA} + I_{DD}} \tag{3.13}$$

When using PIE, the intensity in the acceptor channel after acceptor excitation (I_{AA}) is available as well, which is required to calculate the apparent labeling stoichiometry:

$$S_{app} = \frac{I_{DA} + I_{DD}}{I_{DA} + I_{DD} + I_{AA}} \tag{3.14}$$

Molecules that carry a FRET donor but no acceptor dye, adopt a stoichiometry value of ~ 1, while the stoichiometry of acceptor-only molecules is close to zero. Intermediate stoichiometry values are thus associated with double-labeled molecules. The different species can be vizualized by plotting the labeling stoichiometry against the proximity ratio as shown in Figure 3.5A. Besides the acceptor- and donor-only populations (indicated by red and green boxes, respectively), two double-labeled species (highlighted by a yellow box) are observed, one at a low-FRET and one at a high-FRET efficiency. The trailing between the single- and double-labeled molecules indicates photobleaching and -blinking. As only photostable, double-labeled molecules are of interest for the FRET analysis, single-color species as well as photobleaching and -blinking events have to be removed. To this end, the alternating laser excitation two-channel kernel-density estimator (ALEX-2CDE) filter (Tomov et al., 2012) was applied, which estimates photon densities around each detected photon and is sensitive to brightness fluctuations. Figure 3.5B shows the distribution of the ALEX-2CDE filter plotted against the labeling stoichiometry. Acceptor- and donor-only molecules show high values for the filter, while the double-labeled molecules appear at a value around 10. The photophysics-related trailing between the populations exhibits values > 10. Thus, an upper limit of 10 for the ALEX-2CDE filter removes donor- and acceptor-only molecules and most of the subpopulation mixing. In Figure 3.5C, only the double-labeled molecules and their apparent FRET efficiencies are shown. For a quantitative analysis, several correction parameters need to be considered, which can be determined directly from each measurement (Hellenkamp et al., 2018; Kudryavtsev et al., 2012).

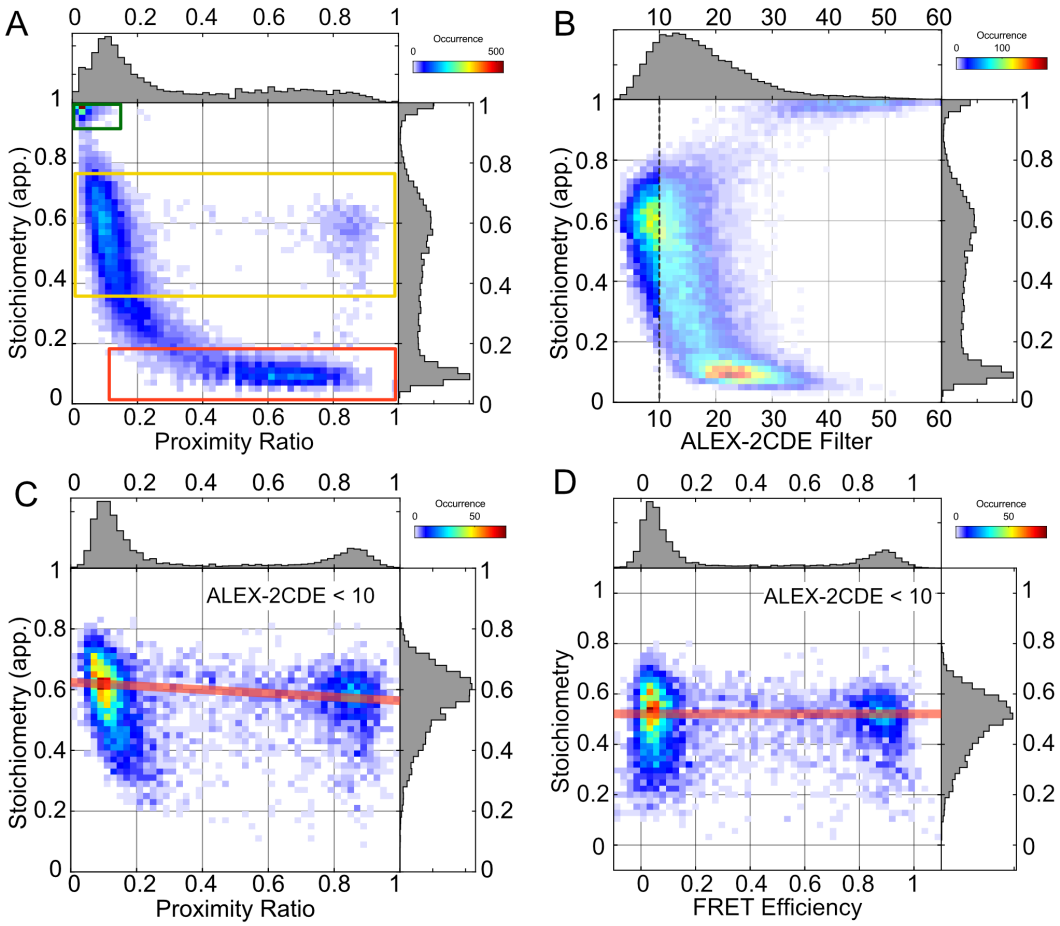


FIGURE 3.5: Typical workflow for the selection and correction of solution-based smFRET data, based on a measurement of a dual-labeled RNA stem-loop (dyes: Atto532/647N) in the presence of the RNA chaperone NSP2 (chapter 5). **A** A two-dimensional histogram of apparent FRET efficiencies (proximity ratio) versus apparent stoichiometry of all molecules selected by the all-photon burst search algorithm. Donor-only, double-labeled and acceptor-only molecules are indicated by green, yellow, and red boxes, respectively. **B** A plot of the ALEX-2CDE filter against the apparent stoichiometry helps to identify the upper threshold of the filter to remove donor- and acceptor-only molecules as well as trailing between the populations caused by photobleaching and -blinking. **C** A two-dimensional histogram of the proximity ratio versus apparent stoichiometry for all molecules exhibiting an ALEX-2CDE value < 10. As the data are not γ-corrected, the two FRET populations adopt different stoichiometry values, as highlighted by the red line. **D** A two-dimensional histogram of corrected FRET efficiencies versus stoichiometry for all molecules exhibiting an ALEX-2CDE value < 10. Both FRET populations adopt a stoichiometry value of $S = 0.5$.

The spectral crosstalk of the donor fluorescence into the acceptor detection channel α can be obtained from the donor-only population:

$$\alpha = \frac{E_{PR}^{D_{\text{only}}}}{1 - E_{PR}^{D_{\text{only}}}} \tag{3.15}$$

The direct excitation of the acceptor fluorophore by the donor laser can be calculated from the acceptor-only species:

$$\delta = \frac{S_{app}^{A_{\text{only}}}}{1 - S_{app}^{A_{\text{only}}}} \tag{3.16}$$

Finally, differences in detection efficiencies η and quantum yields ϕ of the two dyes are accounted for by the γ-factor, which is given by

$$\gamma = \frac{\eta_A \phi_A}{\eta_D \phi_D} \tag{3.17}$$

The γ-factor can be calculated by minimizing the differences in stoichiometry of two or more FRET species, or by using the lifetime information of the donor fluorophore (Kudryavtsev et al., 2012). Considering the correction factors, the accurate FRET efficiency E and labeling stoichiometry S are thus defined by

$$E = \frac{I_{DA} - \alpha I_{DD} - \delta I_{AA}}{I_{DA} - \alpha I_{DD} - \delta I_{AA} + \gamma I_{DD}} \tag{3.18}$$

and

$$S = \frac{I_{DA} - \alpha I_{DD} - \delta I_{AA} + \gamma I_{DD}}{I_{DA} - \alpha I_{DD} - \delta I_{AA} + \gamma I_{DD} + \beta I_{AA}} \tag{3.19}$$

The additional factor β in Equation 3.19 considers differences in cross-sections σ (or extinction coefficents ϵ) and excitation intensities I of the two fluorophores:

$$\beta = \frac{\sigma_A I_A}{\sigma_D I_D} \tag{3.20}$$

Using the β-factor results in $S = 1/2$ for double-labeled molecules with a 1:1 ratio of donor to acceptor dyes, as shown in the corrected FRET efficiency versus stoichiometry plot in Figure 3.5D.

Fluorescence lifetime and anisotroy

As the confocal microscopes used in this work combine PIE with MFD, several fluorescence parameters can be obtained simultaneously from every experiment, including the lifetime and anisotropy of the fluorophores. In *PAM*, the calculation of the molecule-wise fluorescence lifetime is based on a maximum likelihood estimation (Kudryavtsev et al., 2012; Maus et al., 2001). The ideal fluorescence decay is convoluted with the instrument response function of the system (determined from a clean, non-fluorescent PBS measurement) and background scattering (extracted from a buffer measurement) added. This model is then compared to the actual data and the lifetime varied to minimize the deviation. As statistics in burst experiments are limited, lifetimes are determined by fitting the fluorescence decay to a single-exponential model.
The donor lifetime τ_D is directly linked to the FRET efficiency and the relation for a static molecule given by

$$E_{\text{static}} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \tag{3.21}$$

with the lifetime $\tau_{D(A)}$ and $\tau_{D(0)}$ of the donor in the presence and absence of the acceptor, respectively. Equation 3.21 defines the static-FRET line (shown as a red solid line in Figure 3.6A), which is slightly curved due to the attachement of the dyes by flexible linkers. Even for static samples, this linker flexibility causes fast fluctuations in the inter-dye distance. As the lifetime fit is based on the donor fluorescence decay only, low FRET efficiency states (i.e. states with higher donor signal) contribute more photons to the decay and thus shift the curve towards longer lifetimes. Whenever a molecule undergoes dynamic conversions while diffusing through the confocal volume, it adopts different FRET efficiencies, which are averaged per burst. As τ_D is calculated from the signal intensity of the donor, it is biased towards the brighter, low-FRET efficiency species. A rightward shift from the static-FRET line is thus observed for dynamic molecules. Dynamic-FRET lines (Kalinin et al., 2010) can be defined between two states with different FRET efficiencies and donor lifetimes τ_1 and τ_2:

$$E_{\text{dyn}} = 1 - \frac{\tau_1 \tau_2}{\tau_{D(0)}(\tau_1 + \tau_2 - \langle \tau \rangle_f)} \tag{3.22}$$

where the observable $\langle \tau \rangle_f$ is the fluorescence-weighted average lifetime. As a deviation from the static-FRET line is observed in Figure 3.6A, two dynamic-FRET lines are additionally defined to describe the data. As most of the events lie on the dynamic-FRET line between the FRET efficiency states at $E = 0.9$ and $E = 0$ (dashed blue line) and not on the dynamic line between the states at $E = 0.9$ and $E = 0.5$ (dashed red line), blinking of the donor fluorophore is indicated.
The acceptor lifetime τ_A does not depend on the FRET efficiency, but gives valuable information on potential acceptor quenching, which reports on the quantum yield of the dye. In case of varying acceptor lifetimes, different γ-factors would be required for data correction. The exemplarly plot of τ_A versus FRET efficiency

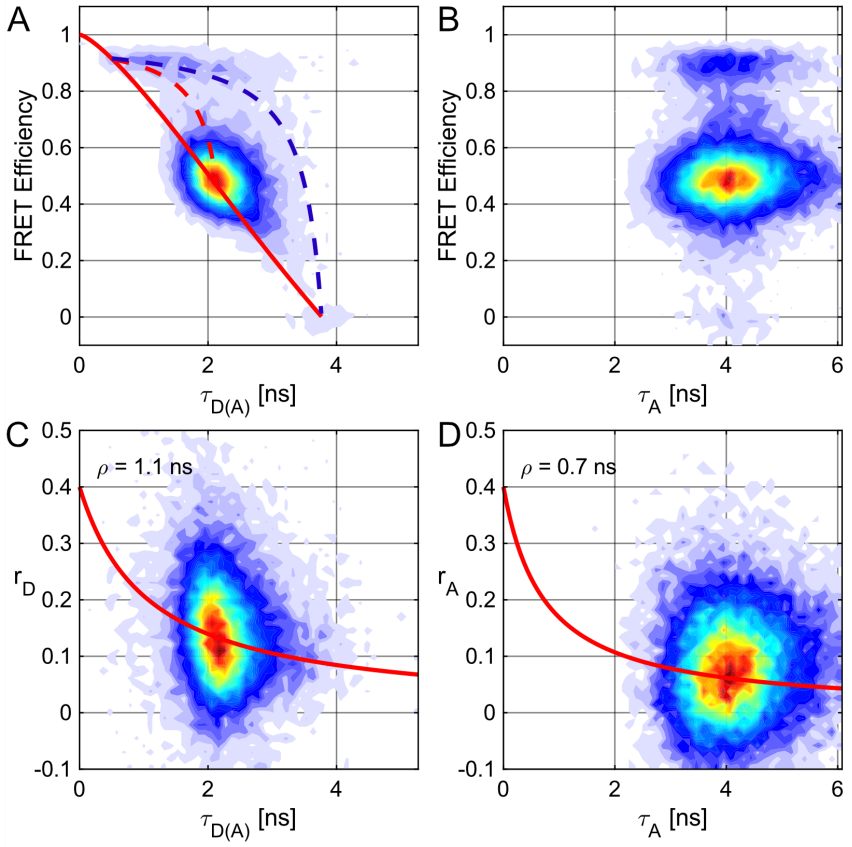


FIGURE 3.6: Histograms of fluorescence lifetime and anisotropy obtained from a PIE-MFD experiment on the dual-labeled protein SdrG 548-595 in the presence of 1.25 μM unlabeled target peptide Fgβ (project 3). Donor- and acceptor-only molecules were removed with the ALEX-2CDE filter (ALEX-2CDE < 8). **A** Two-dimensional histograms of the FRET efficiency versus the donor lifetime. The static-FRET line is shown as a solid red line and the dynamic-FRET lines as dashed red and blue lines. **B** Two-dimensional histograms of the FRET efficiency versus the acceptor lifetime with acceptor excitation. The FRET efficiency does not depend on the acceptor lifetime. **C** and **D** Two-dimensional histograms of fluorescence anisotropy and lifetime of the donor and acceptor fluorophore, respectively. The rotational correlation time ρ is calculated with the Perrin equation (indicated by a red line).

in Figure 3.6B indicates two main populations at $E = 0.5$ and $E = 0.9$ and a minor population at $E = 0$, all exhibiting identical lifetimes $\tau_A = 4\,\text{ns}$. As no acceptor quenching is observed, the same γ-factor can be assumed for the data set.
The fluorescence anisotropy is connected to the rotational mobility of the dyes and can give information on the interaction with larger binding partners or sticking of the fluorophores. In the case of a single-exponential decay of the intensity $I(t) = I_{\parallel} + 2I_{\perp}$, the relation between the steady-state anisotropy r (see Equation 2.4) and the fluorescence lifetime τ is given by the Perrin equation:

$$r = \frac{\int_0^\infty r(t)I(t)dt}{\int_0^\infty I(t)dt} = \frac{r_0}{1 + \frac{\tau}{\rho}} \tag{3.23}$$

with the fundamental anisotropy of the fluorophore r_0 and the rotational correlation time ρ. Two-dimensional histograms of anisotropy versus lifetime of donor and acceptor dyes are given in Figure 3.6C and D, respectively. In both cases, one main population is detected. Multiple species with identical correlation times would fall on the same Perrin line.

Burst-wise FCS

A fluorescence correlation analysis cannot only be performed at nM concentrations by investigating temporal signal fluctuations on the ensemble level as described in Section 3.1.4, but also burst-wise at pM concentrations for indiviual FRET populations. Differences in diffusion time can be detected by a species-selective FCS analysis, which was first introduced by Eggeling and co-workers (Eggeling et al., 1998). When the correlation time is limited to the duration of the fluorescence burst, the approach strongly depends on the chosen burst search parameters. A low selection threshold will allow the detection of longer bursts and diffusion times, while a higher selection threshold will cut the bursts and underestimate the diffusion time. To render the analysis independent of the burst duration, a time window can be added before and after each burst (Laurence et al., 2007). If a second burst is detected during the time window of a first one, both are excluded from the analysis. All other bursts will contribute to a species-selective autocorrelation curve. A burst-wise FCS analysis was performed for the NSP2 project in chapter 5, to confirm the binding of the RNA chaperone protein to the low- and high-FRET populations of a dual-labeled RNA stem-loop.

Photon distribution analysis (PDA)

To extract accurate distances from smFRET experiments, a photon distribution analysis (PDA) (Antonik et al., 2006) was performed.
The number of photons detected in either the donor or FRET channel during a fluorescence burst is limited. Due to this photon shot-noise, discrete FRET efficiency states appear not as defined peaks, but as broad distributions, even if the molecules adopt only a single conformation. Additionally, the presence of different conformational substates will also influence the width of the distribution. To unambiguously distinguish the shot-noise contribution from physically relevant heterogeneities in the FRET experiment, PDA targets the uncorrected FRET efficiency (or proximity ratio) histogram and calculates its shot-noise limited, minimum width based on stochastic processes and the experimental correction factors including spectral crosstalk, direct excitation of the acceptor fluorophore by the donor laser and effects of the background. Thus, underlying distance variations can be determined.
An exemplarly PDA fit of the protein SdrG, labeled stochastically with the dye pair Atto532 and Atto643 at positions 277 and 595, is shown in Figure 3.7. The spikes and voids in the histogram at e.g. $E_{RP} = 0.5$ and $E_{RP} = 0.67$ are caused by photon shot-noise and accounted for by PDA. The main population (in orange) is centered around a distance value of 54 Å with a distribution width of 2 Å. A second population was included in the fit (shown in blue) to describe the additional low-FRET efficiency population caused by dye blinking..
These experimental distances can then be compared to theoretical values based on AV calculations (Kalinin et al., 2012) on crystal structures or computer simulations to validate the smFRET approach or to define quantitative structural models (Dimura et al., 2016), as discussed in chapter 6.

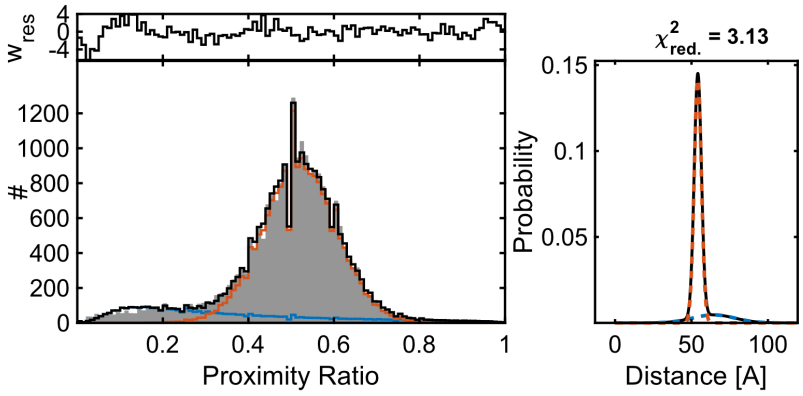


FIGURE 3.7: Exemplarly photon distribution analysis (PDA) of the protein SdrG, which is labeled with Atto532 and Atto643 at positions 277 and 595, using time bins of 1 ms for the procximity ratio histogram. The main population (in orange) is fitted by a center distance of 54 Å and a distribution width of 2 Å.

3.2 Single molecules on the surface

The diffusion time of a molecule limits its observation time in solution-based experiments. Even relatively large molecular assemblies such as the DNA origami platform discussed in chapter 4 do not spend more than 30 ms on average in the confocal detection volume. To observe individual molecules over a longer time range, they have to be either tracked (Ruthardt et al., 2011) or immobilized on a surface. In this thesis, surface-anchored proteins and nucleic acids were measured on a total internal reflection fluorescence (TIRF) microscope.

3.2.1 Total internal reflection fluorescence microscopy (TIRFM)

When light encounters an interface between a medium of high refractive index (n_2) and a medium of low refractive index (n_1), for example at the interface between a glass coverslip and an aqueous buffer, it is either refracted into the second medium or reflected at the interface, depending on the incident angle θ_2 (Figure 3.8). For small angles θ_2, the light is refracted under angle θ_1 and the relation between the refractive indices and angles is given by Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{3.24}$$

At a sufficiently high incident angle, termed critical angle θ_c, θ_1 reaches a value of 90° and the refraction direction thus becomes parallel to the interface. The critical angle is calculated as:

$$\theta_c = \sin^{-1}\frac{n_1}{n_2} \tag{3.25}$$

For incident angles above θ_c, the beam is totally internal reflected within the high refractive index medium. However, the incident beam can penetrate a short distance (in the range of 100 nm) into the low refractive index medium as an evanescent wavefront in the near field. The intensity I of this evanescent field decreases exponentially with distance z according to

$$I(z) = I_0 \exp\left(-\frac{z}{d}\right) \tag{3.26}$$

where I_0 is the intensity directly at the interface and the decay constant d defined by

$$d = \frac{\lambda_0}{4\pi\sqrt{{n_2}^2 \sin\theta_2 - {n_1}^2}} \tag{3.27}$$

where λ_0 is the wavelength of the incident light in vaccum. As the excitation volume is restricted to the evanescent field, only fluorophores located close to the interface are excited.

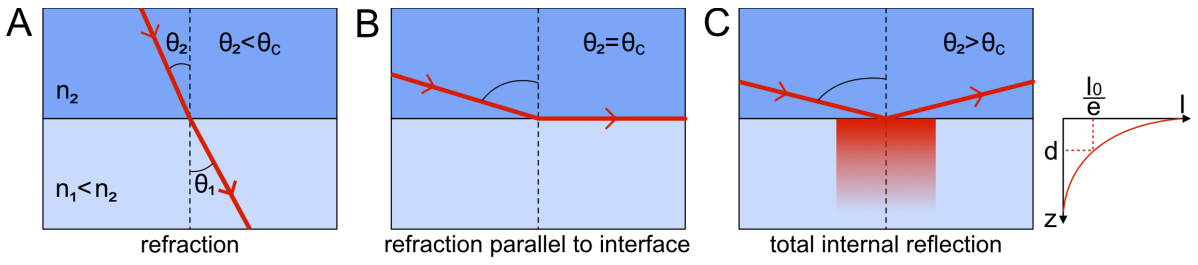


FIGURE 3.8: Light at the interface of two media with different refractive indices n_1 and n_2, with $n_2 > n_1$. Depending on the incident angle θ_2, light is either refracted (**A**, **B**) or reflected (**C**).

Two main excitation configurations exist to generate an evanescent field: prism-type and objective-type TIRF, illustrated in Figure 3.9. For prism-type TIRF, the incoming laser beam is directed into a quartz prism and the incident angle adjusted beyond the critical angle. The total internal reflection of the beam generates an evanescent field below the prism where the sample is located and fluorescence emission is collected by a water immersion objective. For objective-type TIRF, the excitation beam is focused off-axis on the back focal plane of a high numerical aperture oil immersion objective (NA of 1.4 or higher) to realize an incident angle above the critical angle. The evanescent field is generated on top of the glass coverslip and emission is collected by the same objective. For this thesis, mostly objective-type TIRF experiments were performed, as this configuration

allows an easy access and manipulation of the sample. The experimental setup is described in detail in the following section.

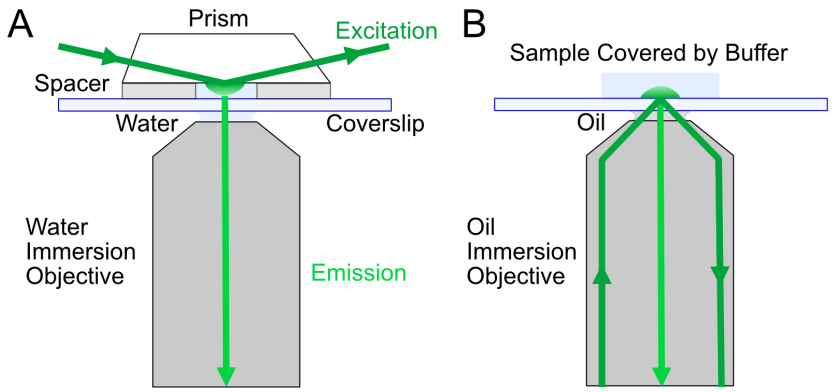


FIGURE 3.9: The two main excitation configurations for TIRF microscopy. **A** In prism-type TIRF, the excitation beam is reflected on the interface between a quartz prism and the sample buffer. **B** For objective-type TIRF, incident angles above the critical angle are realized by shifting the excitation beam off-axis on the back focal plane of a high numerical aperture objective.

3.2.2 Objective-type TIRF setup

Two color-smFRET experiments of surface-immobilized biomolecules were performed on a home-built TIRF microscope, which offers both prism- and objective-type excitation. The microscope was constantly modified over the course of this thesis including a move of the whole setup to another building. A schematic of the objective-type configuration, that was used for most experiments presented in this thesis is shown in Figure 3.10.
For excitation, laser lines at 647 nm (Cobolt MLD, 120 mW, Solna, Sweden) and 561 nm (Cobolt Jive, 50 mW) or 532 nm (Cobolt Samba, 100 mW) are combined through dichroic mirrors into a single-mode fiber (optimized for 532/647 nm, OZ Optics, Carp, Canada) using an apochromatic laser coupler (60SMS-1-4-RGBV11-47, Schäfter+Kirchhoff, Hamburg, Germany). An acousto-optical tunable filter (AOTFnC.400-650-PV-TN, Pegasus Optik, Wallenhorst, Germany) allows the intensity and duration of the laser lines to be selected for millisecond alternating laser excitation (ALEX, (Kapanidis et al., 2005)). At the fiber exit, the laser beam is collimated by an apochromatic collimator (60FC-4-RGBV11-47, Schäfter+Kirchhoff) and focused onto the back focal plane of a 60x oil immersion objective (Apo TIRF 60x/1.49 Oil, Nikon, Düsseldorf, Germany).

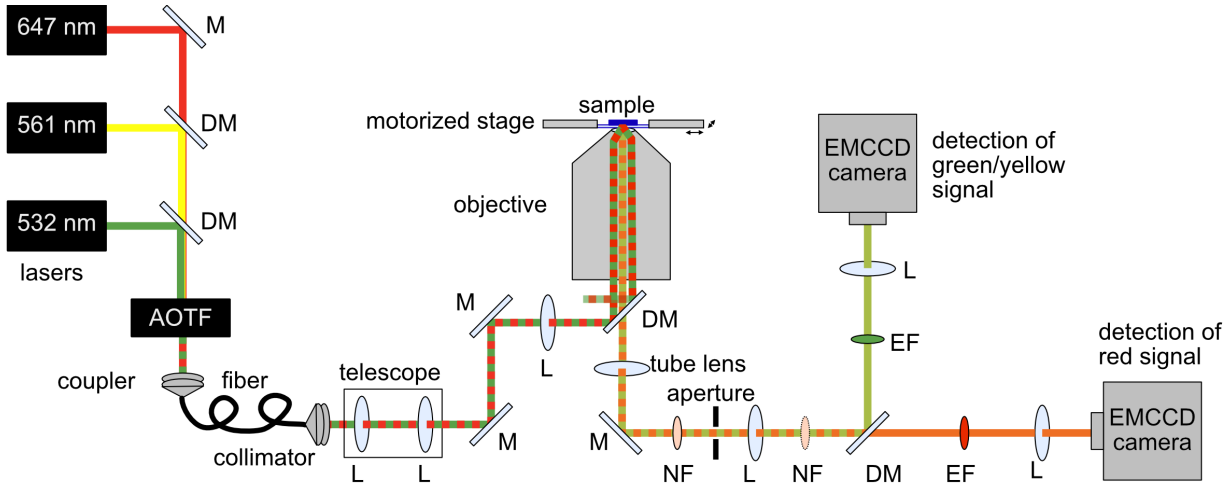


FIGURE 3.10: Schematic representation of the objective-TIRF setup used in this work. M: Mirror, DM: Dichroic Mirror, L: Lens, NF: Notch Filter, and EF: Emission Filter.

Fluorescence is collected by the same objective and seperated by a dichroic mirror (z532/633rpc, AHF Analysentechnik, Düsseldorf, Germany) or a polychroic mirror (405/488/561/647, 630DCXR, AHF Analysentechnik), depending on the excitation wavelengths chosen. The fluorescence signal is focused by the tube lens onto a rectangular aperture, which is located in the image plane just outside of the side port of the microscope body, and then collimated by a second achromatic lens. Scattered excitation light is removed by a 488/647 notch filter (AHF Analysentechnik) and a switchable 532/10 notch filter (AHF Analysentechnik) in case the 532 laser is used. The fluorescence signal is split by a dichroic mirror (630DXCR, AHF Analysentechnik), transmitting red emission and reflecting green and yellow emission. The red emission is cleaned by a HQ 705/100 filter (AHF Analysentechnik) and focused by an achromatic lens onto an Electron Multiplying Charge Coupled Device (EMCCD) camera (Andor iXON, Andor Technology, Belfast, UK). Green and yellow emissions are cleaned by a HQ 595/50 or HQ 580/75 HQ filter (both AHF Analysentechnik), respectively, which can be easily exchanged using a filter wheel. An achromatic lens focuses the signal on the EMCCD camera (Andor iXON) used for green or yellow detection. Both cameras and the AOTF are connected to a field-programmable gated array (FPGA, cRIO-9073, National Instruments, Austin, Texas, USA) for synchronized triggering and acquisition. The control software was written by Bässem Salem and Frank Mieskes in Labview (National Instruments, Austin, Texas, USA).

3.2.3 Preparation of flow cells for surface immobilization

TIRF experiments were performed in flow cells, which allow the specific immobilization of molecules carrying a biotin-tag to a biotinylated surface via streptavidin, as well as an easy exchange of buffer solutions during the experiment.
The preparation procedure is based on a protocol provided by Evelyn Plötz. To prepare a functionalized glass surface, microscope slides (borosilicate, 24 mm × 60 mm, thickness 170 ± 5 µm, Paul Marienfeld GmbH & Co. KG, Lauda-Königshofen, Germany) were first sonicated for 20 min in HPLC-grade acetone. The slides were rinsed three times with millipore water, sonicated for 20 min in HPLC-grade ethanol, rinsed again with millipore water and dried with a stream of compressed air. To remove any organic material left on the surface, the glass slides were plasma-cleaned for 15 min with oxygen. The slides were then incubated in 50 ml toluene containing 300 µl polyethylene glycol(PEG)-silane (6-9 PEG units, abcr GmbH, Karlsruhe, Germany) and 5 mg biotin-PEG silane (Nanocs Inc., NY, USA)[1] for at least 12 hours at 55 °C. Coating the glass surface with PEG prevents unspecific binding and the low percentage of biotin-functionalized PEG provides specific anchor points for streptavidin, which can then bind the biotinylated molecules of interest. After incubation, the glass slides were rinsed two times with absolute ethanol, sonicated for 5 min in absolute ethanol, rinsed with millipore water and dried with a stream of compressed air. If immediately used to assemble flow cells, the slides were photobleached for 10 min using a UV lamp (254 nm and 366 nm) to reduce the fluoresence background. Any additional slides were stored under vacuum at room temperature for a maximum of 14 days.
Besides the biotin-PEG functionalized microscope slide, a small quadratic glass slide (20 mm × 20 mm) with two holes drilled through it and two triangular pieces of parafilm were required to assemble a flow cell (Figure 3.11).

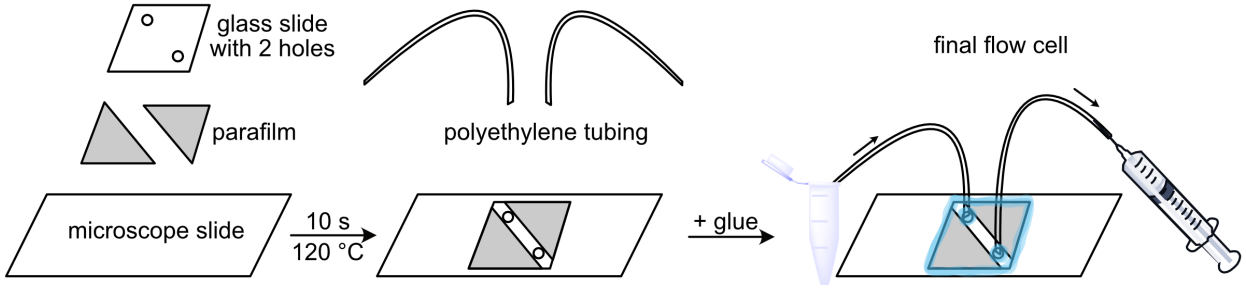


FIGURE 3.11: Preparation of a flow cell for TIRF experiments. Required components: a biotin-PEG functionalized microscope slide, parafilm spacers and a glass slide with two holes for in- and outlet tubings. The flow cell is sealed with glue and connected to a syringe to generate a flow of buffer or sample solution.

A sandwich of these components was placed on a heating plate for about 10 seconds at 120 °C to melt the parafilm, which tightly connected the small glass slide to the microscope slide and created a channel between

[1] Biotin-PEG silane, which was ordered as a powder, was dissolved in water-free solvent (DMSO or chloroform), aliquoted and stored at -80° C.

the two holes. Excessive melting of the parafilm was avoided to prevent clogging of the channel. After cooling, excess parafilm around the small glass slide was removed with a cutter. Two roughly 15 cm long pieces of polyethylene tubing (Smiths Medical Portex Fine Bore LDPE Tubing, inner diameter 0.58 mm, outer diameter 0.96 mm) were then plugged in the holes. To fix the tubings in place and seal the flow cell, a two-component epoxy glue was mixed and applied around the holes and the small glass slide with a clean pipette tip. After 30 minutes, the glue was sufficiently dried and the flow cell ready to use.
For the measurements, the flow cell was mounted on the movable stage on top of the objective and fixed with adhesive tape. The two tubings were connected to a syringe or placed into an eppendorf tube with the buffer or sample solution, creating out- and inlets, respectively. By pulling the syringe, liquid was transferred from the eppendorf tube into the flow cell. As a first step, the flow cell was usually flushed with PBS to check for any leakage from the channel and dirt contamination on the glass surface. After an incubation step with streptavidin, the biotinylated molecules of interest were added to the flow cells.
The small glass slides with the holes were re-used and had to be cleaned after every experiment. To disassemble the flow cells, they were incubated in acetone for at least three days. Remaining glue on the small glass slide was then removed using tissues soaked in acetone. Special attention was paid to the inlet and outlet holes, which were scratched free using a needle. The pre-cleaned slides were sonicated for 20 min in 2 % Hellmanex, rinsed three times with millipore water, sonicated for 20 min in HPLC-grade acetone, rinsed three times with millipore water, sonicated for 20 min in absolute ethanol, rinsed three times with millipore water, sonicated for 20 min in 1 M KOH, rinsed three times with millipore water and dried with a stream of compressed air.

3.2.4 SmFRET experiments using alternating laser excitation

As already mentioned in section 3.1.3, it is advantageous to not only excite the donor, but also the acceptor fluorohore directly in smFRET experiments. Instead of pulsed lasers, continuous wave lasers and fast shutters can be used for an alternating laser excitation (ALEX). The concept of ALEX on different time scales was originally introduced by Kapanidis et al., 2005. For the experiments presented in this thesis, lasers were alternated on the millisecond timescale using an AOTF controlled by a FPGA. Typical exposure times were 30 ms or 100 ms, followed by a frame transfer of the EMCCD camera of 3 ms. An exposure time of 30 ms thus resulted in a total frame time of 66 ms for a two-color FRET experiment. Although the time resolution is reduced compared to a single-color excitation scheme, the use of ALEX is strongly adviced to distinguish real FRET transitions from apparent dynamics caused by photophysical artifacts, as illustrated in Figure 3.12.

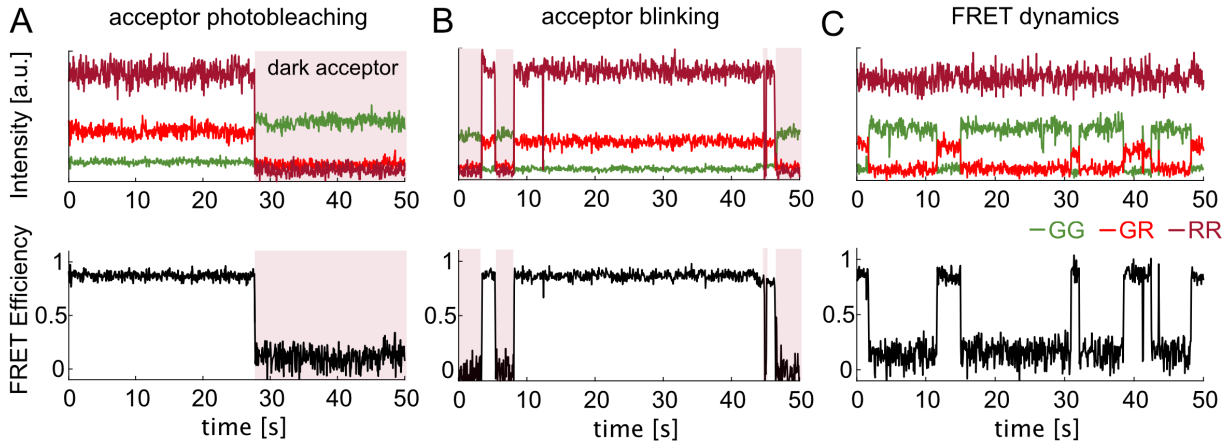


FIGURE 3.12: ALEX helps to distinguish photophysical artifacts such as acceptor photobleaching (**A**) and blinking (**B**) from dynamic FRET signals (**C**) in smTIRF experiments. The data shown are based on a measurement of surface-immobilized DNA origami platforms with 6-nt overhang DNA double-strands (chapter 4) with a frame time of 66 ms. The green signal after green excitation (GG) is shown in green, the red signal after green excitation (GR, i.e. FRET signal) is shown in red, and the red signal after red excitation (RR) is shown in dark red.

Shown are exemplary fluorescence intensity time traces with the donor signal after donor excitation in green, the acceptor signal after donor excitation (i.e. the FRET signal) in red and the acceptor signal after acceptor excitation in dark red. In Figure 3.12A, the acceptor and FRET intensities drop after 28 s and the intensity of the donor dye increases. Without the acceptor information, the anti-correlated FRET and donor signals could be misinterpreted as a dynamic transition from a high-FRET to a low-FRET efficiency state. However, with ALEX, this event is unambigiously identified as acceptor photobleaching. In Figure 3.12B, the acceptor

cycles between a fluorescent and a dark state. In its photoactive state, the acceptor can receive energy from the donor dye, leading to a high-FRET efficiency signal. When the acceptor enters the dark state, the FRET signal decreases and the donor signal increases, as it is no longer quenched by FRET. This causes anti-correlated FRET and donor signals, which are, however, not due to FRET dynamics but caused by acceptor blinking. Please note that these anti-correlated signals are observed because the acceptor entered a non-accepting triplet state here. For dark states caused by the transfer of an electron and thus the formation of an radical ion, the acceptor would still be able to accept energy from the donor dye albeit no FRET signal would be detected. Only Figure 3.12C shows real FRET transitions between a high-FRET and a low-FRET efficieny state, as the acceptor signal after acceptor excitation remains stable over time. The ALEX information is thus crucial for choosing appropriate traces for analysis. The workflow for the extraction and analysis of smTIRF data is described in the following section.

3.2.5 Data analysis

To analyse the experimentally recorded smTIRF data, two MATLAB-based programs were used: *TRacer* written by Bässem C. Salem, for camera mapping and intensity trace extraction, and *Trace Intensity Analysis (Tracy)*, originally developed by Gregor Heiss and modified by Evelyn Plötz and Anders Barth, for further analysis.

Camera mapping

As the emission signals of the donor and acceptor fluorophores for every FRET pair were spectrally seperated and detected on two different cameras, the same molecules had to be first localized in both channels. To this end, a transformation map was generated to find the corresponding pixels on both cameras, correcting for optical abberation and an imperfect alignment of the cameras. The pattern of evenly spaced holes in the aluminium film of a zero-mode waveguide (ZMW) (Levene et al., 2003) served as a calibration reference. At the beginning of every measurement day, a short movie of a ZMW device illuminated by the microscope lamp was recorded. The positions of the individual apertures were identified in both channels and mapped onto each other (Figure 3.13). This transformation map was then applied to the experimental data to find FRET pairs across both channels. For further analysis, only double-labeled molecules were considered that showed fluorescence intensity after both donor and acceptor excitation.[2]

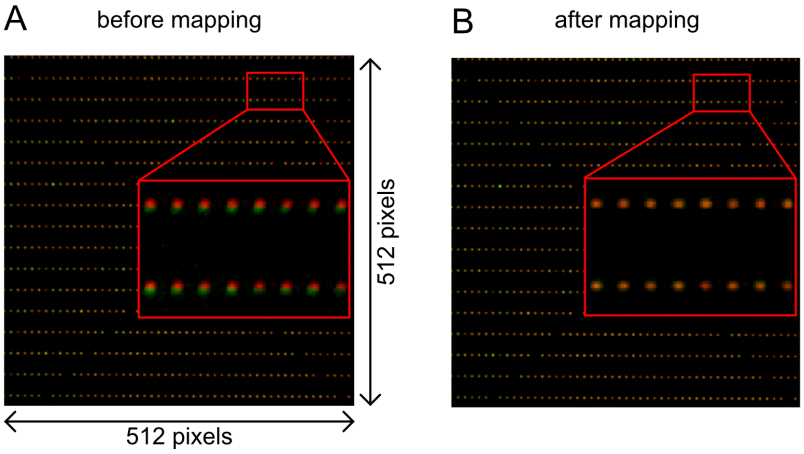


FIGURE 3.13: A transformation map based on the pattern of apertures on a ZMW device. The signal detected on the green EMCCD camera is shown in green and the signal detected on the red camera is shown in red. **A** Before mapping, the signals in the two channels do not overlap perfectly due to optical aberrations, especially in the upper part of the camera chip (highlighted by the red box). **B** The overlap is improved after the transformation map is applied. Overlapping signals of the green and red cameras are shown in yellow.

[2]Using *TRacer*, the particles found in both detection channels are numbered consecutively (column-wise from left to right) to mark them for intensity trace extraction. When the same field of view is imaged repeatedly, this feature is extremely useful to identify the corresponding time traces of individual molecules in the different recorded movies, as shown in the multi-enzyme TIRF experiment presented in chapter 4.

Intensity extraction and calculation of correction factors

After localizing the individual molecules in the two channels, frame-wise intensities after donor and acceptor excitation were extracted inside a circular signal mask covering the PSF of the molecule. To correct for local background noise, a ring-shaped mask around the signal mask was defined. The intensitiy within this ring was averaged over five frames and the median value substracted from the intensity within the signal mask. Background-corrected intensity traces of the donor signal after donor excitation (I_{DD}), acceptor signal after acceptor excitation (I_{AA}) and acceptor signal after donor excitation (I_{DA}) were then exported to *Tracy* for further analysis.
The time traces were scanned for photobleaching steps and automatically sorted into the categories 'no photobleaching' 'acceptor photobleached', 'donor photobleached' and 'both photobleached'. Traces with an acceptor photobleaching step before donor photobleaching (Figure 3.14) were used to calculate two correction factors for accurate FRET efficiencies: the spectral crosstalk of the donor emission into the acceptor detection channel, α, and the detection correction factor, γ. The spectral crosstalk was calculated as the ratio of FRET-sensitized acceptor signal to donor intensity after acceptor photobleaching

$$\alpha = \frac{\langle I_{DA}^{A,\text{bleached}} \rangle}{\langle I_{DD}^{A,\text{bleached}} \rangle} \tag{3.28}$$

where $\langle I_{DA}^{A,\text{bleached}} \rangle$ and $\langle I_{DD}^{A,\text{bleached}} \rangle$ denote the mean FRET and donor intensities after acceptor photobleaching, respectively. As mentioned before for the confocal measurements (see Equation 3.17), γ accounts for differences in detection efficiencies of the cameras and quantum yields of donor and acceptor fluorophores, and can be experimentally determined as the ratio of FRET and donor intensity changes after acceptor photobleaching

$$\gamma = \frac{\langle I_{DA} \rangle - \langle I_{DA}^{A,\text{bleached}} \rangle}{\langle I_{DD} \rangle - \langle I_{DD}^{A,\text{bleached}} \rangle} = \frac{\Delta I_{DA}}{\Delta I_{DD}} \tag{3.29}$$

The photobleaching step of the donor is in principle not required to calculate α and γ, as only the donor intensity after acceptor photobleaching or relative changes in donor intensities before and after acceptor photobleaching are considered. However, the single photobleaching step confirms the presence of only one donor and excludes a contribution of additional fluorophores in close proximity to the FRET pair of interest (either due to diffusion of unbound fluorescent molecules or clustering of immobilized molecules on the surface). Using *TRacer*, ΔI_{DA} can be additionally corrected for direct excitation of the acceptor fluorophore by the donor laser, and for spectral crosstalk of the donor emission into the acceptor detection channel (Salem et al., 2019). It is further possible to calculate the γ-factor from dynamic FRET traces without photobleaching steps by minimizing the variance of the total intensity (sum of FRET and donor signals) as a function of γ. However, as the presence of multiple dyes can not be excluded, this approach was not used to correct the data presented in this thesis.

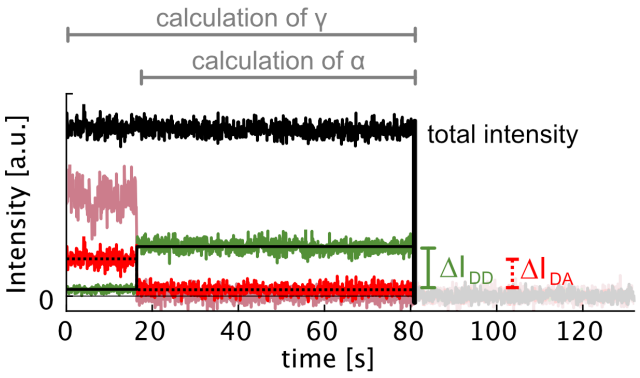


FIGURE 3.14: Calculation of correction factors for smFRET TIRF experiments based on an intensity time trace with an acceptor photobleaching step (after 18 s) before donor photobleaching (around 80s). The mean intensities of the donor and FRET signals before and after acceptor photobleaching are indicated by a solid and dashed black line, respectively. The relative intensity changes are used to calculate the γ-factor (Equation 3.29) and the intensities after acceptor photobleaching are used to calculate the spectral crosstalk α (Equation3.28). The corrected total intensity (in black) is increased by a factor of 2.5 for better visibility and shows no fluctuations before the donor is bleached.

Due to the use of photostabilizing agents, usually only a few traces exhibited photobleaching of both fluorophores. The correction factors determined from these traces were then averaged and the mean value applied to the rest of the traces. Using the correction factors, framewise FRET efficiencies were calculated for every molecule

$$E = \frac{I_{DA} - \alpha I_{DD}}{I_{DA} - \alpha I_{DD} + \gamma I_{DD}} \tag{3.30}$$

Trace selection criteria

Next, the traces were manually sorted into different categories. Built-in filters in *Tracy* helped to pre-select molecules according to their averaged FRET efficiency value and to identify potential dynamic FRET traces by scanning for anti-correlated donor and FRET signals. To qualify for selection, every trace had to fullfill several requirements:

- Both FRET dyes had to be photoactive and the emission after acceptor and donor excitation stable over time. Traces showing blinking, fast photobleaching of one or both dyes, or a gradual decrease in intensity (caused by sample drift or defocusing) were excluded from further analysis.
- Ideally, traces should show one donor and/or one acceptor photobleaching step to determine correction factors and to confirm the presence of only one donor and one acceptor dye. Traces with several photobleaching steps were excluded from further analysis.
- If traces did not show photobleaching, the signal intensities were compared to traces with one donor and/or acceptor photobleaching step. Traces with higher intensities than the ones expected for a single dye pair were excluded from further analysis.

Having passed this quality control, the molecules were sorted according to their FRET signal into either static or dynamic traces. A molecule was considered to exhibit a static FRET signal, when the intensity in the acceptor channel after donor excitation stayed constant until one of the fluorophores bleached. An exemplary static FRET trace with an acceptor photobleaching step is shown in Figure 3.12A. A dynamic molecule fluctuating between different FRET efficiency states was identified by anti-correlated FRET and donor signals, while the acceptor intensity after acceptor excitation remained constant (Figure 3.12C). However, FRET transitions faster than the frame time (in the case of a typical ALEX experiment: 66 ms) or slower than the measurement time (132 s for 2000 frames) cannot be resolved using TIRF.

Analysis of dynamic FRET traces by a hidden Markov Model (HMM) approach

For molecules fluctuating between several FRET states, transition rates were calculated using a hidden Markov model (HMM, McKinney et al., 2006; Rabiner, 1989; Zarrabi et al., 2018). In a Markov model, transitions between discrete observable states are described by a stochastic process without memory termed a Markov process. Thus, the transition probability from one state to another depends only on the present state. In a hidden Markov model, which was originally developed for speech regognition (Baum and Petrie, 1966), the states are not directly observable. In the case of smFRET TIRF data, discrete FRET efficieny states are hidden within the experimental noise, resulting in a distribution of FRET efficiency values which can be approximated by a Gaussian emission function with mean value μ and standard deviation σ. A HMM is based on two functions: The probability distribution function, which is a combination of all emission functions describing the measured observables, and the transition probability function, which gives the probability for transitions between two hidden states. An iterative expectation-maximization approach (Baum et al., 1970) was used to determine the optimal parameters to describe the data. Using this information, the most likely kinetic sequence of states, the Viterbi path (Viterbi, 1967), can be determined for an individual trace.
As initial input parameters, the number of FRET states and their approximate FRET efficiency value (μ of the emission function) were estimated from framewise FRET efficiency histograms, which include the FRET efficiency values of all selected molecules determined at every frame of the movie. An exemplary framewise histogram is shown in Figure 3.15A. As the histogram displays two main population, two FRET states are assumed for the HMM. Hidden parameters were then iteratively optimized until a defined threshold for the change in the log-likelihood was reached (the convergence threshold was usually set to 10^{-6}). Two main approaches exist to train HMM on smFRET time traces: either emission parameters and transition probabilities are calculated independently for every single molecule (local HMM) or one set of values is used to fit all molecules (global HMM). The data presented in this thesis were fitted by a local HMM to better account for

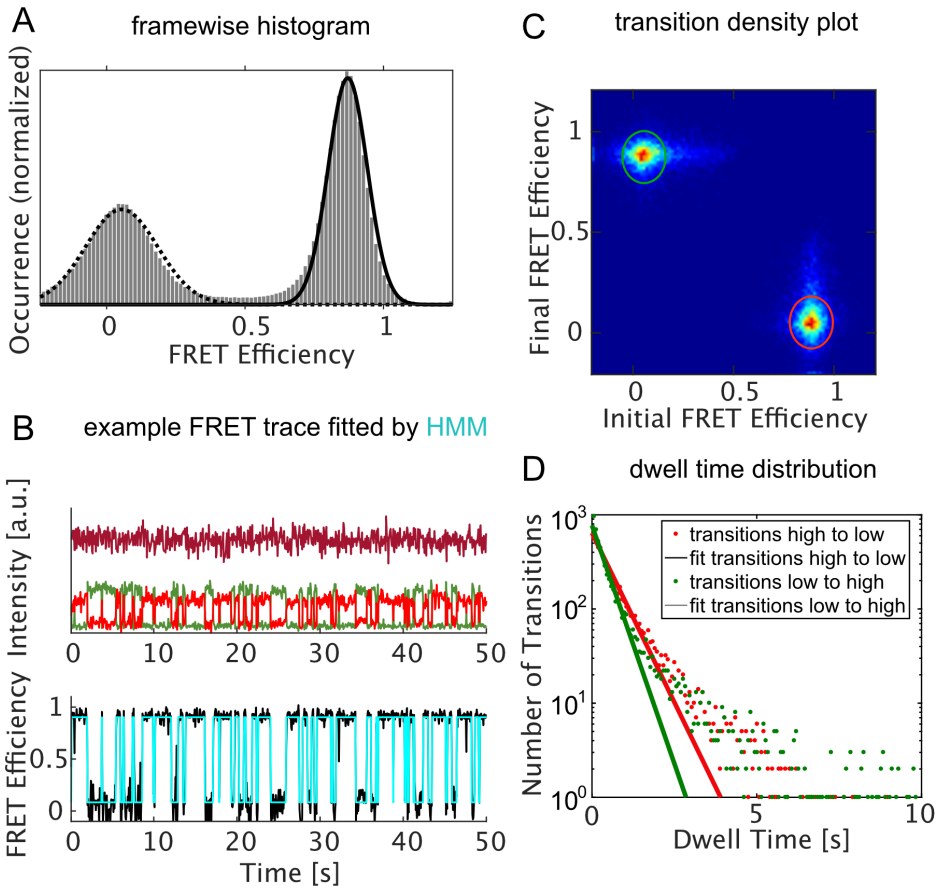


FIGURE 3.15: Illustration of the workflow to analyze dynamic smFRET TIRF data. The data shown are based on a measurement of surface-immobilized DNA origami platforms with 6-nt overhang DNA double-strands (chapter 4) recorded with alternating green-red laser excitation over 2000 frames with an inter-frame time of 66 ms. The total number of molecules is 169. **A** The framewise histogram of FRET efficiencies can be described by two Gaussian distributions (shown as dashed and solid black lines), indicating two different FRET states. **B** Exemplary intensity time trace of a dynamic molecule with anticorrelated donor (green) and FRET signals (red). The acceptor signal after acceptor excitation (dark red) remains constant over time. The FRET efficieny (lower graph, black) fluctuates between a low-FRET and a high-FRET state and is well approximated by the Viterbi path determined from a local HMM analysis (cyan). **C** The corresponding transition density plot illustrates the switching between low- and high-FRET efficiencies. **D** The dwell time histograms of the clusters marked in C (in red and green, respectively) with mono-exponential fits (red and green lines, respectively). The mean dwell time in the high-FRET state is 641 ms, and the mean dwell time in the low-FRET state 770 ms.

molecule-by-molecule heterogeneities. Figure 3.15B shows the Viterbi path (in cyan) of an exemplary FRET efficieny trace, that was fitted by a local HMM. To vizualize the results of a local HMM model applied to all selected time traces, FRET efficieny values before and after a transition were plotted in a transition density plot (TDP, Figure 3.15C). In a TDP, the connectivity of different states is directly displayed and similar transitions can be identified as clusters. By selecting a cluster of transitions, the respective dwell time in the initial state can be fitted by a mono-exponential decay, as HMM assumes constant probabilities and thus exponential rates, and the transition rate calculated as the inverse of the dwell time. The mono-exponential fits presented in Figure 3.15D describe the fast transitions (i.e. with short dwell times) well. To describe the few slower transitions, a second exponential fit is required, which indicated degenerated FRET efficiency states.

Chapter 4

A DNA Origami Platform for Single-Pair Förster Resonance Energy Transfer Investigation of DNA-DNA Interactions and Ligation

In this project, DNA nanotechnology is combined with spFRET to study transient interactions of DNA strands, their ligation and subsequent digestion by a restriction enzyme. Due to the modularity of DNA origami, the platform is ideally suited to study a wide range of multistep reactions and weak interactions. This work has been published in *JACS* and the article with a detailed description of the experimental procedures and the results is given in the appendix.This chapter summarizes the key results and provides additional information on the preparation of DNA origami samples for single-molecule experiments.

A DNA Origami Platform for Single-Pair Förster Resonance Energy Transfer Investigation of DNA-DNA Interactions and Ligation by Kira Bartnik, Anders Barth, Mauricio Pilo-Pais, Alvaro H. Crevenna, Tim Liedl, and Don C. Lamb *J. Am. Chem. Soc.* 142, 815-825 (2020)

4.1 Motivation

DNA is a highly stable, complex organic material that stores our genetic information. However, even under normal cellular conditions, it is susceptible to spontaneous changes and damages. Composed of two strands forming a double helix, DNA is ideally suited for repair, as there are two seperate copies of the genetic material. When both strands of the double helix break, however, a repair template is not easily available. As these DNA double-strand breaks (DSBs) can lead to fragmentation or rearrangement of chromosomes when not properly repaired, they are considered to be the most cytotoxic form of DNA damage (Aparicio et al., 2014). Daily, each cell aquires up to 10^5 spontaneous DNA lesions (Hoeijmakers, 2009), among them around 10 DSBs (Lieber, 2010). These can be a result of normal cellular processes or ionizing radiation. Chemotherapeutics can cause additional DSBs (Limp-Foster and Kelley, 2000). In humans, non-homologous end-joining (NHEJ) (Lieber, 2010) and homologous recombination (HR) (San Filippo et al., 2008) are the major repair pathways to restore genomic integrity. HR relies on an intact sister chromatid as a template for repair and is thus restricted to the S/G2 phase, while NHEJ is active throughout the cell cycle.
Repair by NHEJ is the main pathway in higher organisms and extensive studies have been performed to understand its underlying processes. In several *in vitro* experiments, damaged plasmid DNA was incubated with cell extracts as a source of repair components (Pastwa et al., 2009). In the last years, the kinetics of the NHEJ repair process were mostly investigated by live-cell imaging of fluorescently-tagged NHEJ components or immunofluorescence (Abdisalaam et al., 2014; Britton et al., 2013; Mari et al., 2006; Yano and Chen, 2008). In a recent *in vitro* study, recombinant NHEJ proteins were used to repair fluorescently-labeled double-stranded DNAs (dsDNAs) and the mechanism was investigated by single-pair Förster Resonance Energy Transfer (spFRET) (Reid et al., 2015). New insights into the kinetics of the end-joining process could be gained and a two-step mechanism of DNA end-pairing followed by an alignment within the synaptic complex was proposed. However, the experimental approach relied on the random pairing of a freely diffusing dsDNA with a surface-immobilized dsDNA, which does not depict the close proximity of broken DNA ends in the cell.

To this end, I designed a DNA origami platform to specifically mimic a DNA DSB in a defined geometry. Due to the induced proximity on the origami, the interaction and end-joining of the two DNA double strands can be investigated on the single-molecule level.

4.2 Introduction of the method: DNA origami combined with single-molecule FRET

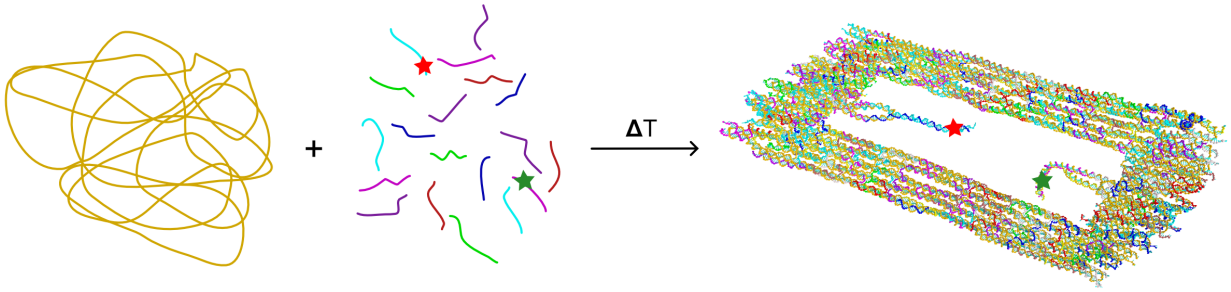


FIGURE 4.1: Schematic assembly of a DNA origami. The scaffold strand (in gold) and a mixture of short staple strands (in various colors) fold into the desired DNA origami object. The example shown is the two-layer DNA origami frame that was used to study DNA-DNA interactions and ligation. By using fluorescently-labeled staple strands, fluorophores (indicated by red and green stars) can be precisely positioned on the nanostructure.

The term *DNA origami* describes the self-assembly of DNA into arbitrary two- and three-dimensional structures (Douglas et al., 2009; Rothemund, 2006). A DNA origami is composed of a set of short DNA single strands, called "staples", which fold a long, usually circular single-stranded DNA, referred to as the "scaffold", into the desired shape (Figure 4.1). Experimentally, the scaffold strand is mixed with an excess of staple strands (usually at a 5- to 10-fold molar excess), heated and then slowly cooled down to enable the formation of complementary Watson-Crick base pairs. To design a DNA origami nanostructure, the open-source software CaDNAno developed by Douglas et al., 2009 is available. Based on either a honeycomb or square lattice, mono- and multi-layered objects can be designed by suitably routing the scaffold strand. The scaffold is typically derived from the M13 bacteriophage genome (e.g. the p7560 scaffold used in this project) and dictates the sequence of the staple strands. Only recently, the design and production of custom scaffold sequences have been demonstrated as well (Engelhardt et al., 2019).
DNA origami nanostructures allow the precise positioning of (bio)molecules on the nanometer (Saccà et al., 2010; Stein et al., 2011) and even sub-nanometer scale (Funke and Dietz, 2016; Hartl et al., 2018). This makes them ideal platforms for single-molecule studies as a high local concentration of the molecules of interest can be realized while the overall concentration is maintained low enough for single-molecule detection. By placing fluorescent dyes on definded positions (as indicated in Figure 4.1), single-molecule FRET experiments become possible on the DNA origami breadboard. The combination of DNA nanotechnology with FRET has been applied to create advanced biosensors (Selnihhin et al., 2018) or artifical light harvesting systems (Nicoli et al., 2017), study force-dependent DNA-protein interactions (Nickels et al., 2016) or understand the working principles of DNA nanomachines (Khara et al., 2018; Kopperger et al., 2018).
Here, we use a DNA origami platform to locally increase the concentration of two fluorescently-labeled DNA repair substrates and measure FRET between the donor and acceptor dye. Figure 4.2 shows a schematic of the two-layer DNA origami frame with attachment sites for the DNA double strands. When the DNA double strands move within the origami gap, the distance between the dyes changes, which influence the efficiency of the energy transfer. Thus, the FRET efficiency can be used as a readout to distinguish if the two DNA strands are far apart (no FRET), dynamically bind and unbind (dynamic FRET) or are successfully repaired into a single double strand (static FRET).

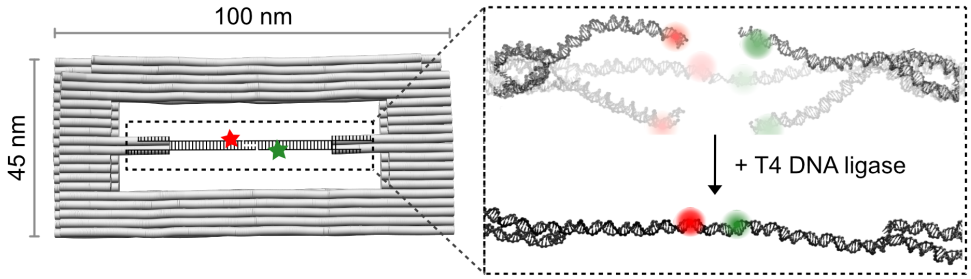


FIGURE 4.2: A DNA origami platform for single-molecule detection of DNA end-joining and ligation. Schematic representation of the DNA origami frame with bound dsDNAs as repair substrates, which are labeled with Atto647N (red star) as FRET acceptor and Cy3b (green star) as FRET donor. The strands can move flexibly before repair and are fixed after repair by the T4 DNA ligase (upper and lower panel in the zoom-in, respectively). Adapted from Bartnik et al., 2020.

4.3 Key results

To catalyze the end-joining of the two dsDNA repair substrates, the DNA ligase from bacteriophage T4 (T4 DNA ligase) was chosen as a model enzyme for repair and the ligation reaction was studied for DNA origamis in solution as well as for surface-immobilized platforms. Due to the modularity of the design, dsDNA repair substrates with different lengths of complementary overhangs and sequences could be bound to the DNA origami and their ligation efficiencies compared.

Freely diffusing DNA origami platforms were monitored in solution using a confocal PIE-MFD microscope (Kudryavtsev et al., 2012; Müller et al., 2005) introduced in section 3.1.2 (Figure 4.3A). Before addition of the T4 DNA ligase, a predominant low-FRET population at a FRET efficiency E around 0 was detected for the various overhang lengths. While this was the only population for the blunt end (0-nt overhang) construct, an additional high-FRET population at a FRET efficiency E around 0.8 appeared for the sticky ends (complementary overhangs of different lengths), which indicates transient binding. The longer the complementary overhang, the higher the population of molecules in the high-FRET and therefore bound state (Figure 4.3B). After incubation with the T4 DNA ligase, a decrease of the low-FRET population and an increase in the high-FRET population was observed for all overhang lengths (Figure 4.3C), indicating successful ligation. As expected, blunt ends were less efficiently ligated compared to the sticky ends, as less high-FRET events were detected. However, the sticky ends also differed slightly in their behavior and 6-nt overhangs were less efficiently ligated compared to the other two sequences.

We further varied the overall dsDNA length and the incubation conditions. Incubation overnight at 16°C resulted in a more pronounced high-FRET population and therefore more ligated dsDNA compared to an incubation at room temperature for 1 h, while shortening or elongation of the DNA double strands to reveal potential geometrical constraints could not increase the ligation efficiency above the already observed level.

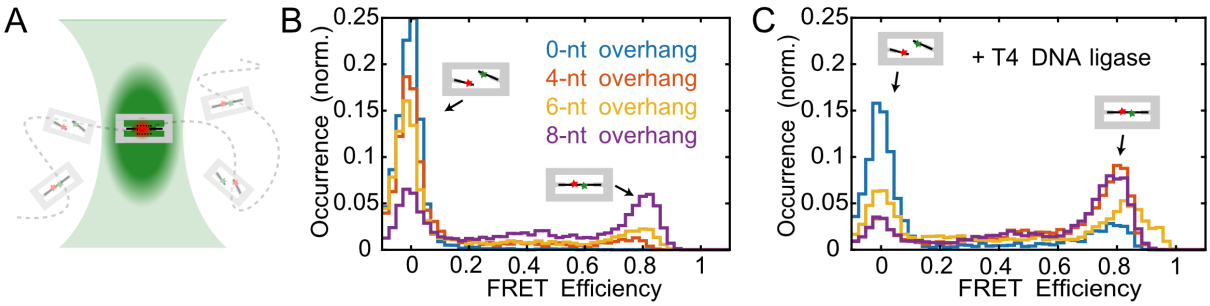


FIGURE 4.3: Detection of DNA ligation in solution. **A** DNA origami platforms with bound fluorescently labeled DNA double strands diffuse through the confocal volume. The red and green stars represent the Atto647N acceptor and Cy3b donor dyes, respectively. **B** Histograms of molecule-wise spFRET efficiencies before the addition of the T4 DNA ligase. The complementary overhang length between the DNA double strands was varied from 0 nt (blue), to 4 nt (orange), 6 nt (yellow) and 8 nt (purple). **C** Histograms of molecule-wise spFRET efficiencies after incubation with the T4 DNA ligase. Adapted from Bartnik et al., 2020.

To follow the ligation of individual dsDNA pairs over the timescale of seconds to minutes, the DNA origamis were immobilized on a glass surface (Figure 4.4A) and measured using TIRF microscopy. For DNA repair substrates with complementary overhangs, transient binding could be observed as dynamic switching between low- and high-FRET states. The kinetic rates were determined by HMM and showed that the dissociation rates are affected by the overhang length between the DNA strands. Using a construct with 4-nt overhangs and a low concentration of the T4 DNA ligase, we were further able to capture the transition from transiently to covalently bound DNA double strands. In the beginning of the FRET efficiency trace in Figure 4.4B, fast fluctuations between a low- and a high-FRET state are observed due to transient binding of the DNA strands. After 30 s, the dynamics stop and a continuous high-FRET signal is detected, indicating a successful ligation reaction.
To test the reversibility of the ligation process, the sequence of the 4-nt sticky-end construct was modified by introducing the recognition site of the restriction enzyme ZraI between the FRET dyes. The DNA origami platforms were immobilzed on a glass surface and first incubated with the T4 DNA ligase, then with ZraI and finally again with the T4 DNA ligase. In Figure 4.4C, anticorrelated donor and acceptor signals are observed before ligation due to transient binding of the sticky overhangs. After incubation with the T4 DNA ligase, a continuous, high-FRET signal indicates a successful end-joining. After incubation with the restriction enzyme ZraI, no FRET is observed, implying digestion of the ligated DNA into blunt ends. After a subsequent incubation with the T4 DNA ligase, the molecule goes back to the high-FRET signal and therefore to the ligated state. Thus, the DNA origami design made it possible to ligate and cut the same DNA double-strand pair repeatedly and the platform stayed intact over several cycles of enzyme treatment.

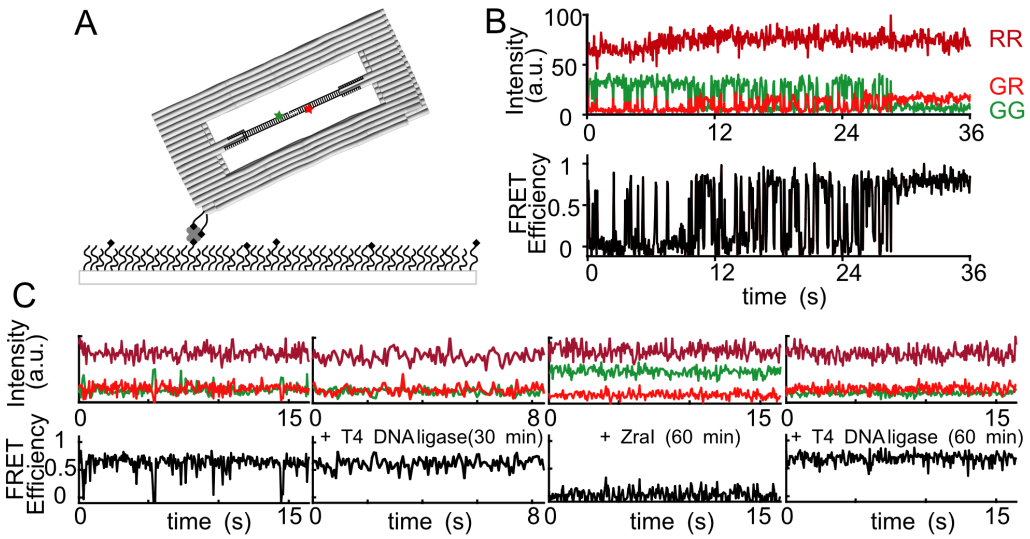


FIGURE 4.4: Detection of DNA ligation on surface-immobilized DNA origamis. **A** A schematic of the DNA origami immobilized on the surface using two biotinylated staple strands. The surface is passivated with PEG and a small fraction of PEG-biotin to anchor the DNA origami via streptavidin for the TIRF measurements. **B** Exemplary FRET trace of origami-bound DNA double-strands with 4-nt overhangs showing the transition from a dynamic to a continuous FRET signal after T4 DNA ligase addition. The donor signal after donor excitation is shown in green, the acceptor signal after donor excitation (i.e., the FRET signal) is shown in red, and the acceptor signal after acceptor excitation (i.e., the ALEX signal) is shown in dark red. **C** Exemplary FRET traces of origami-bound DNA double-strands with restriction enzyme recognition site before the addition of the T4 DNA ligase, after ligation, after subsequent incubation with the restriction enzyme ZraI , and after another ligation step. Adapted from Bartnik et al., 2020.

4.4 Side note: Preparation of DNA origami samples for single-molecule experiments

When preparing DNA origami samples for single-molecule experiments, a thorough purification is required to ensure the removal of misfolded structures, aggregates and excess staple strands, especially when the strands are fluorescently labeled. An overview of the different purification methods available is discussed in detail in Wagenbauer et al., 2017. For the DNA origami platform presented here, a combination of PEG precipitation directly after assembly (Stahl et al., 2014) and gel purification with a subsequent ultrafiltration step after incubation with the fluorescently-labeled DNA double strands was chosen. The physical extraction from the agarose gel made it possible to seperate correctly folded structures from aggregates and unbound DNA strands, while the ultrafiltration step was used to exchange the buffer and to concentrate the sample.

Figure 4.5 shows a representative SYBR Green-stained agarose gel image of DNA origami samples, which were incubated with an excess of fluorescently-labeled DNA double-strands with either blunt ends or 4-nt complementary overhangs before loading.Besides the correctly folded DNA origamis (marked by green boxes in Figure 4.5), which migrate similar to the p7560 scaffold strand on the gel, additional species are visible. The slower migrating bands in the upper region of the gel (red boxes in Figure 4.5) are of higher molecular weight than the desired origami structure and are due to aggregation. The faster migrating bands represent unbound DNA double strands, which were added at a 10-fold molar excess to the origami platform to ensure a good labeling efficiency. To test how aggregation influences the platform, both correctly folded and aggregated DNA origami samples were extracted from the gel and monitored in solution using PIE-MFD. The 2D histograms of stoichiometry versus FRET efficiency in Figure 4.5 indicate different behaviors for the aggregated and correctly folded DNA origami platforms. While the aggregated samples show no high-FRET efficiency population for neither 0-nt nor 4-nt overhangs before ligation (red boxes, solid and dashed lines, respectively), a minor high-FRET population is observed for correctly folded platforms with 4-nt sticky end dsDNAs (green box, dashed line). Thus, aggregation seems to impair transient binding of the 4-nt overhangs to each other. Apparently, the dsDNA repair substrates are less flexible on the aggregated platforms and are thus less accessible for the repair enzyme. Therefore, it is not surprising that incubation with the T4 DNA ligase does not lead to a clear high-FRET efficiency species for both overhang lenghts and therefore only a minor fraction of covalently end-joined DNA ends. As a positive control, the correctly folded DNA origami platforms behave as expected: the high-FRET population increases for both 0-nt and 4-nt overhangs after ligation and is more pronounced for the sticky ends than for the blunt ends. It is worth mentioning that the data presented here are based on a ligation reaction for 1 h at room temperature. Thus, the fraction of ligated DNA strands is lower compared to the data presented in Figure 4.3, which were aquired after an overnight incubation at 16°C. As aggregation has an obvious influence on the DNA origami platform, sample purification is essential to avoid artifacts.

SYBR Green, which was used to vizualize the gel bands under UV light, is a DNA intercalator and stays bound to the origamis after gel extraction. This can additionally influence the integrity of the structure and lead to background noise. For the data presented in the publication (Bartnik et al., 2020), the DNA origami samples were extracted from agarose gels which were not stained with SYBR Green. The weak fluorescent signal of Cy3b-labeled DNA double strands after UV radiation was sufficient to identify the correctly folded DNA band for extraction.

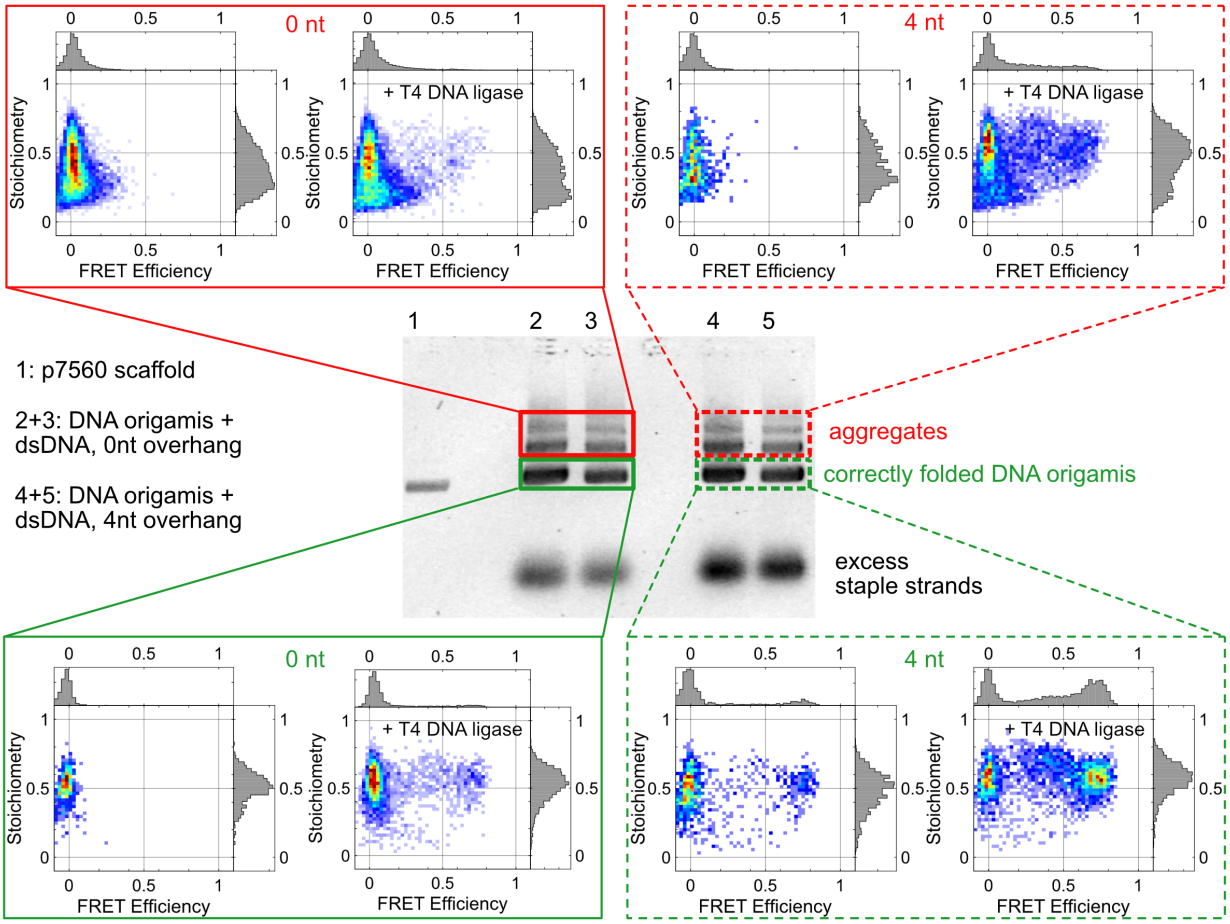


FIGURE 4.5: Purification of DNA origami by gel electrophoresis. **Center** Agarose gel image of the electrophoretically seperated p7560 scaffold strand (lane 1) and DNA origami samples with bound 0-nt or 4-nt overhang dsDNAs (lanes 2 and 3 or lanes 4 and 5, respectively) before incubation with the T4 DNA ligase. The red and green boxes indicate the DNA bands that were recovered from the gel for smFRET measurements. The slow migrating bands are caused by aggregated structures, the fast migrating bands represent excess staple strands. **Upper part** 2D histograms of stoichiometry versus FRET efficiency for DNA origami aggregates before and after incubation with the T4 DNA ligase. **Lower part** 2D histograms of stoichiometry versus FRET efficiency for correctly folded DNA origamis before and after incubation with the T4 DNA ligase for 1 h at room temperature. Solid and dashed lines indicate 0-nt and 4-nt overhang dsDNAs, respectively, while red represents aggregates and green correctly folded DNA platforms.

4.5 Outlook: A versatile platform for interactions on the single-molecule level

The DNA origami platform presented here was used to detect transient binding events of fluorescently-labeled dsDNAs and their covalent end-joining after incubation with the T4 DNA ligase by FRET. However, the origami design is not limited to this specific kind of interaction and can be easily adapted to study interactions between different DNA sequences, proteins and also between DNA and proteins on the single-molecule level, as illustrated in Figure 4.6.

The DNA origami frame is assembled before the molecules of interest are attached to the scaffold protrusions via complementary Watson Crick base pairs. Hence, if a new DNA strand is desired on the platform, the terminal sequence simply has to match the respective scaffold sequence. Thus, various DNA sequences can be bound to the platform and studied in the presence and absence of protein interaction partners (Figure 4.6A and Figure 4.6B, left).

To attach a protein to the origami scaffold, the required complementary DNA sequence can be conjugated via various chemistries. Examples are bifunctional crosslinkers targeting available and reactive amino acid groups like cysteines, lysines, tyrosine, trytophanes or the amino group at the protein's N-terminus, metal affinity

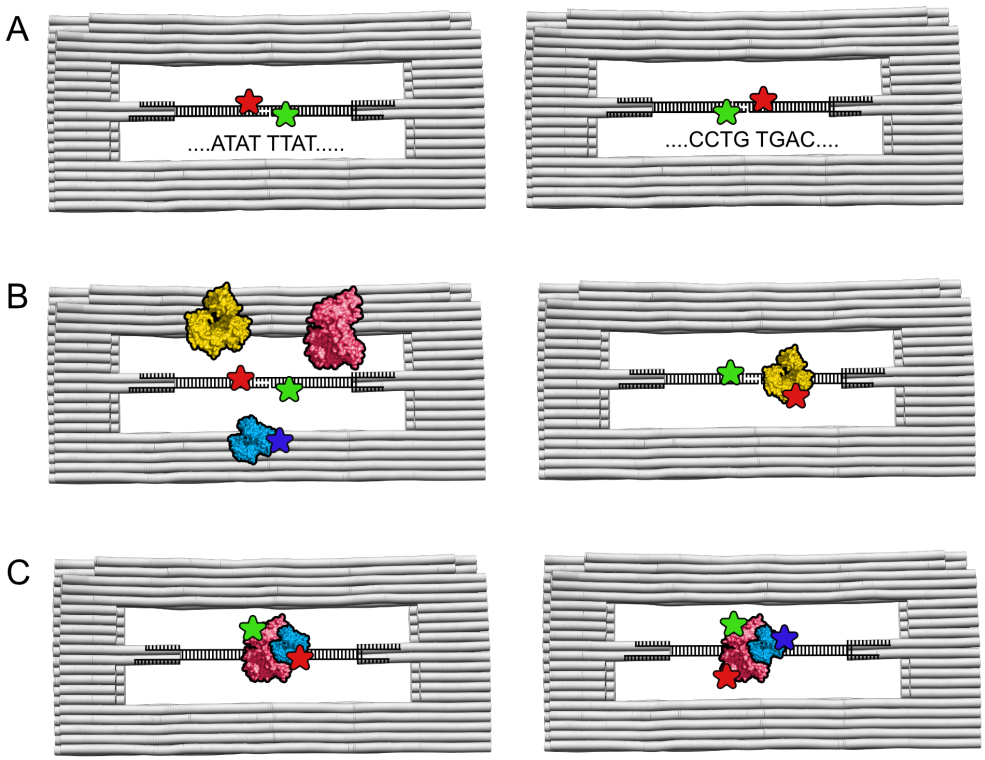


FIGURE 4.6: The DNA origami offers a versatile platform to study biomolecular interactions on the single-molecule level. Shown are schematic representations of DNA-DNA (**A**), DNA-protein (**B**), and protein-protein interactions (**C**).

DNA-templated protein conjugation (Trads et al., 2017) or the fusion of a SNAP-tag to the protein (Hahn et al., 2020). By binding one protein and one DNA strand to the origami (Figure 4.6B, right), their interaction can by studied. To investigate the interaction of two single-labeled proteins, they can both be attached to the platform using the required DNA handles (Figure 4.6C, left). Introducing a third fluorophore on one of the origami-bound proteins additionally allows to follow conformational changes upon binding of the interaction partner (Figure 4.6C, right).

The DNA origami platform allows for specific control of the stoichiometry of the interaction partners and brings them close together in a defined geometry. Thus, weak interactions can be studied, which would normally be difficult to realize on the single-molecule level. Further, by adding more interaction partners in solution, multi-step reactions can be performed on the DNA origami platform.

Chapter 5

Structural basis of rotavirus RNA chaperone displacement and RNA annealing

This project aims to understand how the C-terminal region of rotavirus protein NSP2 regulates its RNA chaperone activity. SmFRET experiments performed by me are combined with cryogenic electron microscopy (cryo-EM) and structural proteomics tools provided by collaboration partners to dissect the role of the C-terminal region. This work has been prepared for publication and the manuscript with a detailed description of the experimental procedures and the results is given in the appendix. In this chapter, the smFRET key results are summarized. Additionally, results from smTIRF measurements are presented and a strategy introduced to reduce RNase contamination in TIRF experiments.

Structural basis of rotavirus RNA chaperone displacement and RNA annealing. by Jack P. K. Bravo, Kira Bartnik, Emma H. Gail, Chen Davidovich, Don C. Lamb, Roman Tuma, Antonio N. Calabrese, and Alexander Borodavka. (2020)

5.1 Motivation and key results

The *Reoviridae* family of multi-segmented double-stranded RNA viruses infects a wide range of hosts from plants to humans (Desselberger, 2014). Rotaviruses, one of the several *Reoviridae* genera, are the most common cause of acute gastroenteritis in young children and infect almost every child around the world by the age of 5, causing hundreds of thousands of deaths every year (Bernstein, 2009). A profound understanding of rotavirus replication is thus required to develop potential therapeutics. The rotavirus genome consists of 11 distinct RNAs, which are all crucial for virus assembly and replication (Borodavka et al., 2018). To faciliate sequence-specific RNA-RNA interactions, rotaviruses encode the RNA chaperone non-structural protein 2 (NSP2). The octameric NSP2 exhibits RNA-bindig and -unwinding activities and promotes duplex formation (Bravo et al., 2018; Taraporewala and Patton, 2001). Its C-terminal region (CTR, residues 295-316) is composed of an ampholytic helix containing positively and negatively charged residues and an unstructured linker, and it has been suggested to play a crucial role in rotavirus replication (Criglar et al., 2014).
To dissect the role of the CTR in the RNA chaperone activity of NSP2, we compared chaperone activities of full-length and C-terminally truncated NSP2 (NSP2-ΔC, omitting residues 295-316). As NSP2-ΔC has been shown to be less efficient in annealing distinct rotavirus RNAs compared to the full-length protein (Borodavka et al., 2017), we hypothesized that such a decreased efficiency of strand annealing is due to a reduced helix-unwinding activity of NSP2-ΔC. To test this hypothesis, the unfolding of a dual-labeled RNA stem-loop was investigated using a smFRET-based assay in solution: As the two dyes Atto532 and Atto647N are in close proximity in the folded RNA configuration, a high-FRET efficiency population is expected. Consequently, unfolding by the RNA chaperone will increase the inter-dye distance and thus decrease the FRET efficiency (Figure 5.1A). In the absence of NSP2, a high-FRET efficiency population could be detected, indicating folded RNA stem-loops (Figure 5.1B, upper panel). In accordance with previous studies (Bravo et al., 2018), addition of full-length NSP2 resulted in two protein-bound RNA populations: a high-FRET folded RNA population and low-FRET unfolded RNA population (Figure 5.1B, middle panel). Protein binding was confirmed by species-selective FCS: The shift of the autocorrelation functions to diffusion times slower than that of the RNA stemloop alone (in yellow) indicated NSP2 binding (Figure 5.1C).

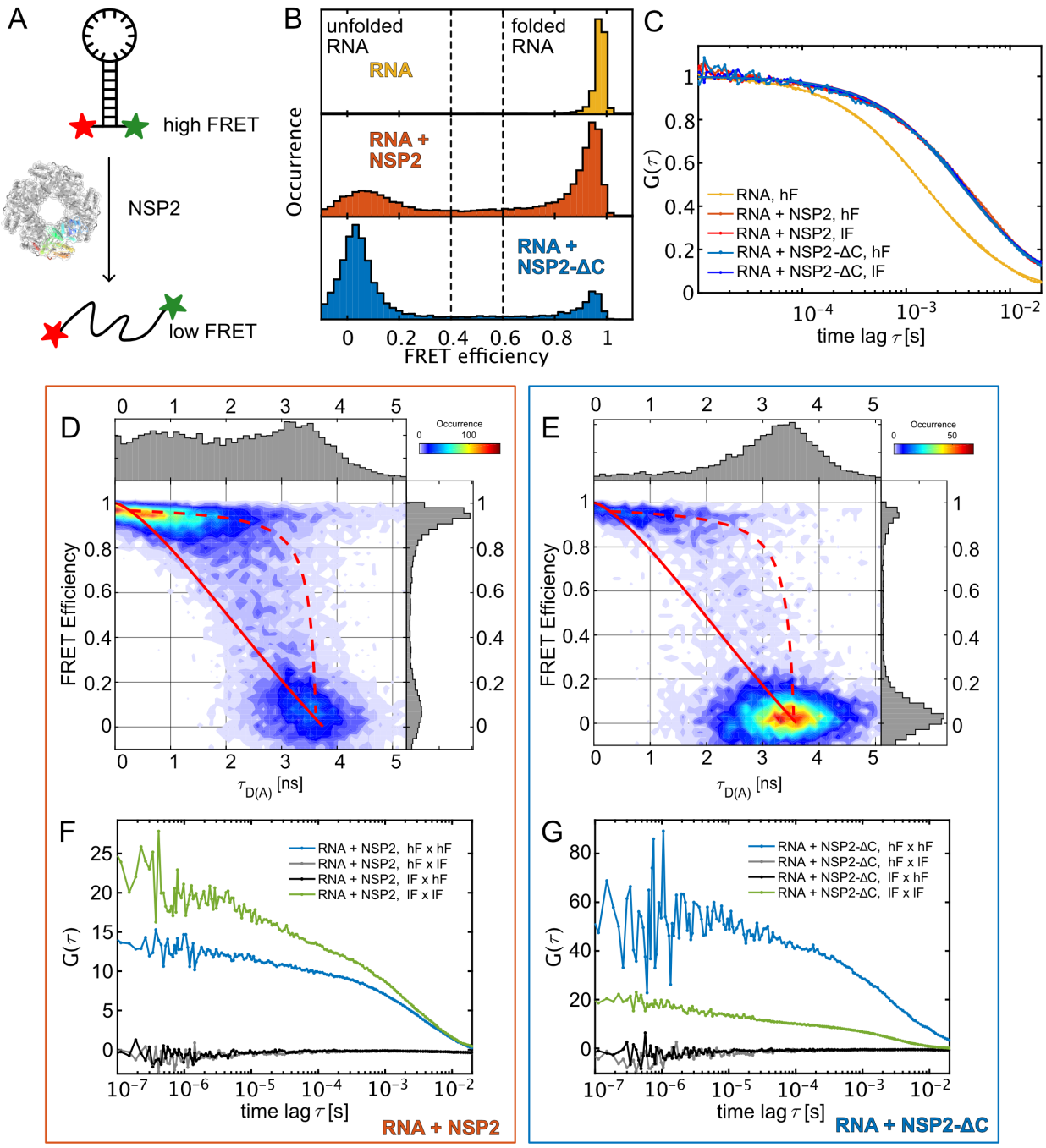


FIGURE 5.1: Enhanced RNA helix unwinding activity of NSP2-ΔC monitored by smFRET. **A** A schematic representation of the unfolding of a dual-labeled RNA stem-loop with Atto532 as FRET donor (green star) and Atto647N as FRET acceptor (red star) by the RNA chaperone NSP2. Shown is a cryo-EM reconstruction of octameric NSP2 with a monomer highlighted in color, adapted from Bravo et al., 2020 (submitted). **B** From top to bottom: smFRET efficiency histograms of the RNA stem-loop (25 pM) alone (yellow) and in the presence of either 5 nM full-length NSP2 (red) or 5 nM NSP2-ΔC (blue). **C** Species-selective FCS analysis of high-FRET (hF) and low-FRET (lF) efficiency populations. The rightward shift of the ACFs from that of RNA alone (in yellow) indicates slower diffusion, and thus binding of NSP2 and NSP2-ΔC for both hF and lF populations. **D-E** Two-dimensional histograms of the FRET efficiency versus the donor lifetime show a deviation from the static FRET line (shown as a solid red line) for RNA in the presence of either full-length NSP2 (D) or NSP2-ΔC (E). Conformational dynamics are expected to follow the dashed red line. **F-G** The species auto-correlation functions (high-FRET in green, low-FRET in blue) and cross-correlation functions (low-FRET x high-FRET in black, high-FRET x low-FRET in grey) extracted by filtered-FCS are shown do not indicate dynamic conversion on the time scale of diffusion for RNA in the presence of full-length NSP2 (F) or NSP2-ΔC (G).

Repeating the same experiment with the C-terminally truncated protein led to a surprising result: While the same two protein-bound low- and high-FRET RNA populations could be observed in the presence of NSP2-ΔC, the equilibrium was clearly shifted towards the low-FRET and thus unfolded fraction (Figure 5.1B, lower panel). Histograms of FRET efficiency versus donor lifetime $\tau_{D(A)}$ in Figure 5.1D and E display a deviation from the static FRET line (solid red line) for RNA in the presence of full-length or C-terminally truncated NSP2, explaining the bridge between low- and high-FRET states observed in the smFRET histograms in Figure 5.1B. However, the shift from the static-FRET line is minor and does not follow the expected dynamic-FRET line (dashed red line). To further investigate potential dynamics on the timescale of diffusion, a filtered-FCS analysis (Barth et al., 2018; Felekyan et al., 2012) was performed, indicating no dynamic conversion between low- and high-FRET RNA populations on the μs-timescale in the presence of full-length NSP2 or NSP2-ΔC (Figure 5.1F and G). The deviation from the static-FRET line might thus be caused by slower dynamics or the binding of multiple RNA stem-loops by the same NSP2 octamer.
Experiments with NSP2 proteins derived from a different Rotavirus strain (RF) confirmed the enhanced RNA helix-unwinding activity of NSP2 in the absence of the CTR, while the RNA annealing was decreased. To resolve these counter-intuitive findings, RNA binding by NSP2 was further investigated using cryoEM, structural proteomics and biochemical assays. The direct characterization of the NSP2-RNA interactions by my collaboration partners revealed that the CTR does not make contacts with the RNA. However, the acidic patch of the CTR was found to be in close proximity to the RNA, indicating a potential role in promoting RNA release via charge repulsion. To investigate if these acidic residues are responsible for the reduced chaperone activity of NSP2-ΔC, a mutant was gernerated lacking only the 6 terminal residues from the unstructured part (residues 311-316). These residue have previously been demonstrated to be non-essential for virus replication (Taraporewala et al., 2006). Indeed, this mutant retained its RNA unwindig and annealing activities comparable to that of the full-length protein. Collectively, these smFRET results strongly suggest that the conserved acidic residues DEED located in the CTR down-regulate the helix-unwinding activity of NSP2 to promote RNA dissociation required for an efficient RNA annealing step.

5.2 Additional single-molecule TIRF experiments

5.2.1 The DNA-RNA hybrid structure for TIRF experiments

To follow the unfolding of individual RNA stem-loops and detect slow dynamics, single-molecule TIRF experiments were performed using a DNA-RNA hybrid structure (presented in Figure 5.2A), which was immobilizated to a streptavidin-coated surface via biotin. For the hybrid structure, the RNA stem-loop sequence used in the smFRET experiments presented above was extended by 20 nucleotides at the 3′-end.[1] The extension allowed the RNA construct to be internally labeled with an Atto647N dye as a FRET acceptor, and, additionally, the installation of a biotinylated terminal nucleotide. The FRET donor was kept at the 5′-terminus, but changed to Cy3b due to the better photophysical properties of this dye for TIRF microscopy. To keep the structure away from the surface, we made it partially double-stranded by annealing a complementary 20-nucleotide DNA sequence,[2] which is not expected to interfere with the RNA chaperone assay, to the RNA. The annealing protocol was the same as for the RNA stem-loop.
The DNA-RNA hybrids were first measured in solution (Figure 5.2B) to confirm their binding and unwinding being similar to the previously used RNA stem-loop (Figure 5.1B). In the absence of NSP2, only a high-FRET efficiency population was observed as expected for the folded RNA stem-loop. As different FRET donor dyes were used for the hybrid and RNA stem-loop and as the FRET dyes were placed farther apart in the hybrid structure, the absolute FRET efficiency values differed slightly. Addition of full-length NSP2 led to low- and high-FRET populations indicating a mixture of folded and unfolded stem-loops, while the C-terminally truncated NSP2 mutant shifted this equilibrium towards the low-FRET fraction. A small, additional high-FRET population was observed, which was attributed to RNA structures missing the complementary DNA strand.

[1] RNA sequence:
5′-(A-Cy3b)AUCUUUGCAAACUAUCCAAUCCAUGACAAAGACGCUAGCUGA(U-Atto647N)GAUCAUCGC-biotin-3′
[2] DNA sequence:
5′-GCGATGATCATCAGCTAGCG-3′

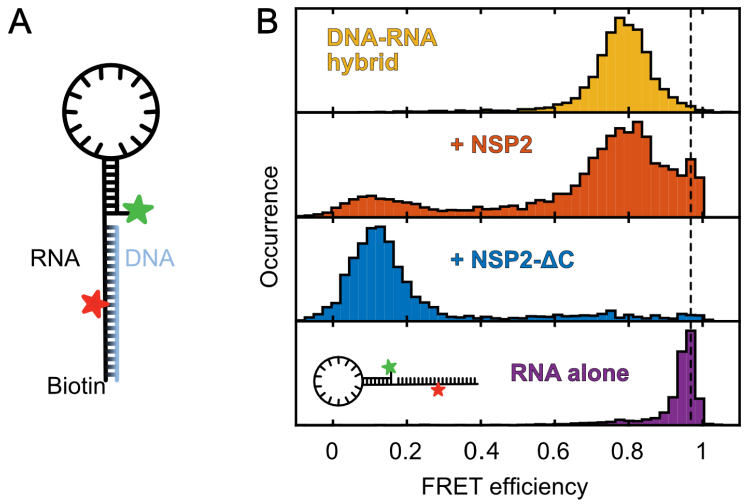


FIGURE 5.2: A DNA-RNA hybrid structure is efficiently unfolded by NSP2-ΔC in solution. **A** Schematic representation of the DNA-RNA hybrid structure with Cy3b as FRET donor (green star) and Atto647N as FRET acceptor (red star). **B** From top to bottom: smFRET efficiency histograms of the RNA-DNA hybrid (50 pM) alone, in the presence of the full-length NSP2 and NSP2-ΔC (100 nM each), and of the RNA stemloop in the absence of the complementary DNA strand. For the latter case, the FRET efficiency of the folded RNA stem-loop is shifted to a higher value, as its single-stranded part collapses in the absence of the complementary DNA strand.

5.2.2 RNase contamination in TIRF experiments

Having established the functionality of the DNA-RNA construct, TIRF experiments were performed with surface-immobilized structures (Figure 5.3A). Unexpectedly, we noticed a substantial loss of the donor signal, which we attribute to contaminating nucleases (RNases) in the used photostabilizing agents. While Trolox (Cordes et al., 2009) was a sufficient photostabilizer for the solution-based measurements presented above, glucose and a glucoseoxidase-catalase (GODCAT) solution were added to surface-immobilized RNA for oxygen removal (Stein et al., 2012). This buffer composition has been used before in TIRF experiments with DNA and resulted in long time traces over seconds to minutes (Bartnik et al., 2020). However, when used with DNA-RNA hybrids, we observed a rapid decrease in the number of molecules displaying a green fluorescence signal after green laser excitation (GG) and a red signal after green excitation (GR, FRET signal), while the red fluorescence after red laser excitation (RR) was not affected (Figure 5.3B).
We reasonend that the position of the green Cy3b dye on the 5'-end made it an easily accessible target for RNase degradation (Figure 5.3A), while the internal labeling position and the heteroduplex protected the red Atto647N dye. To identify whether the glucose solution or the GODCAT mixture was the source of RNase contamination, the cross-correlation function (CCF) amplitudes of freely diffusing DNA-RNA hybrids were measured in the glucose-containing buffer before and after addition of GODCAT (Figure 5.3C). While a high cross-correlation of green and red fluorescence signals could be observed in the glucose buffer, the CCF amplitude dropped to zero upon addition of the enzymatic GODCAT mix. This effect was observed immediately after enzyme addition confirming RNase contamination of GODCAT enzymes. As RNases are much more ubiquitinous than DNases, it is not surprising that DNA in previous experiments was not affected by the photostabilizing agents, while RNA clearly is.
To inhibit RNase activity in TIRF experiments, the flow cell was flushed with 1 ml of a 2 mM vanadyl ribonucleoside complex solution prior to immobilization of the stem-loops. Additionally, the GODCAT solution (10 µl) was treated with 100 U of Murine RNase Inhibitor (2.5 µl of a 40.000 units/ml solution) before mixing with the glucose-containing buffer and adding it to the surface-immobilized DNA-RNA structures. As expected, addition of RNase inhibitor significantly slowed down RNA degradation, permitting detection of donor and FRET signals (Figure 5.3D).

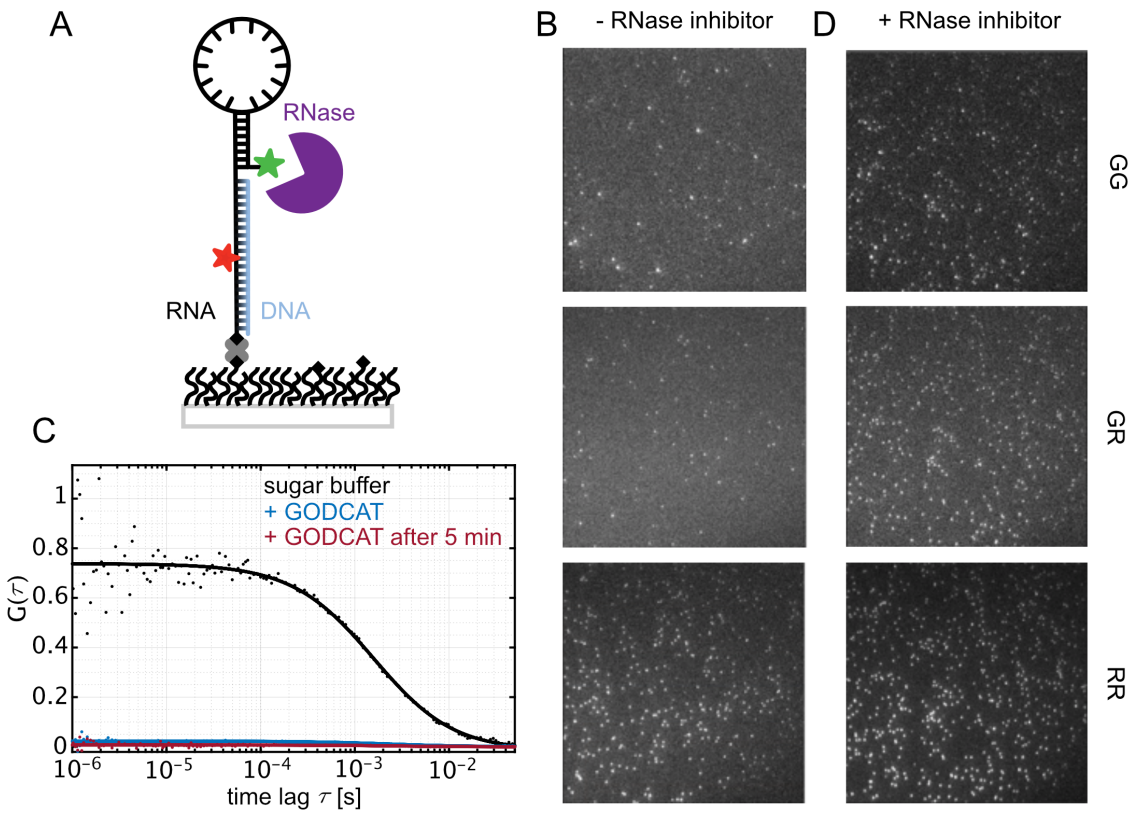


FIGURE 5.3: RNase contamination of commercially available GODCAT enzyme preparations. **A** A schematic representation of the surface-anchored DNA-RNA hybrid. The accessible 5'-end of the RNA labeled with Cy3b is sensitive to RNase degradation. **B** Representative images of surface-immobilized DNA-RNA structures illuminated by TIRF in the absence of an RNase inhibitor. GG: green signal after green excitation, GR: red signal after red excitation i.e. FRET signal, and RR: red signal after red axcitation. **C** Cross-correlation functions of freely-diffusing DNA-RNA hybrids in glucose-buffer and after addition of GODCAT. Rapid loss of cross-correlation upon addition of GODCAT confirms RNA degradation. **D** Same as in B but in the presence of the Murine RNase Inhibitor.

5.2.3 SmTIRF experiments in the presence of RNase inhibitors

TIRF experiments performed in the presence of RNase inhibitors allowed the detection of donor and FRET signals over several seconds to minutes. In the absence of the RNA chaperone, only high-FRET efficiency signals were detected for the DNA-RNA hybrid. The addition of full-length NSP2 or NSP2-ΔC led to a shift to low-FRET. Only a few molecules (usually less than 1 %) displayed dynamic fluctuations between the two FRET states. Thus, statistics were insufficient to reliably calculate the kinetic rates. An example dynamic FRET trace obtained in the presence of 200 nM NSP2-ΔC is presented in Figure 5.4A. The majority of the molecules showed either stable low- or high-FRET signals.
To compare the effects of the full-length NSP2 and NSP2-ΔC, framewise histograms were generated over all recorded FRET traces (i.e. stable high-FRET, stable low-FRET and dynamic FRET traces) (Figure 5.4B). While the DNA-RNA hybrid structures are characterized by a single high-FRET population, incubation with NSP2 results in an additional low-FRET population. The low-FRET state becomes more populated with increasing NSP2 concentration, consistent with RNA helix unfolding. Although NSP2-ΔC slightly increased the low-FRET population compared to the full-length chaperone, the effect is not as pronounced as in the solution-based smFRET measurements (Figure 5.2B). These discrepancies could be attributed to (i) non-specific interactions of the RNA chaperones with the glass surfaces; (ii) different kinetics of RNA unfolding and hybridization close to the surface compared to solution-based assays; (iii) effects of the photostabilization enzyms and/or the RNase inhibitor on RNA binding by NSP2. Due to these limitations, unfortunately we were not able to gain any quantitative insights into the kinetics of RNA unwinding by NSP2 and its C-terminal truncation variant. However, we could demonstrate the absence of intermediate states for RNA stem-loop unfolding. The few dynamic traces we extracted from the experiment further showed transitions on the second timescale, which is indeed to slow to be resolved in diffusion-based measurements.

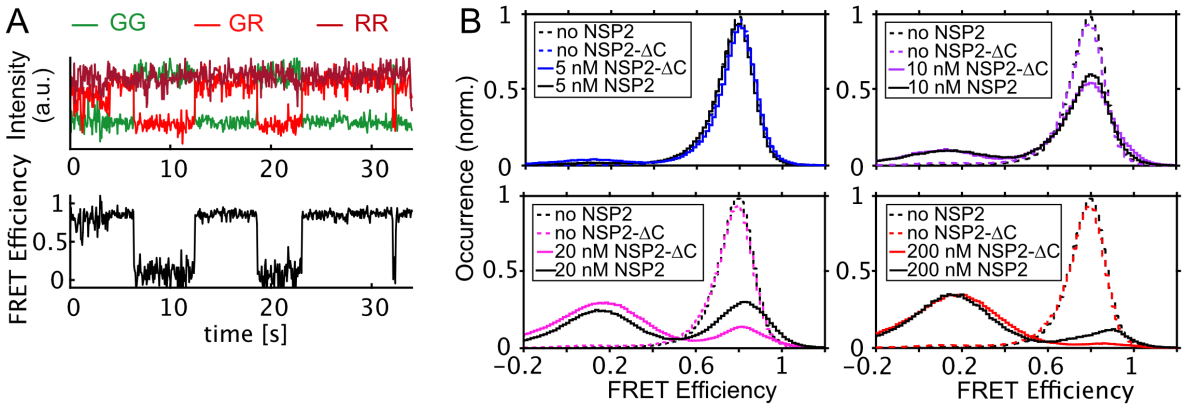


FIGURE 5.4: Surface-immobilized DNA-RNA hybrid structures are unfolded by NSP2. **A** An exemplarly time trace of the DNA-RNA hybrid showing transitions from high- to low-FRET states in the presence of 200 nM NSP2-ΔC. Only a very few such traces were observed. GG: green signal after green, excitation, GR: red signal after green excitation (FRET signal), and RR: red signal after red excitation **B** Framewise histograms over all recorded FRET traces show an increase in the low-FRET population with increasing protein concentration. NSP2-ΔC slightly increases the low-FRET population compared to the full-length chaperone. Data were recorded in the presence of the Murine RNase inhibitor. The numbers of analyzed molecules were 420 in the absence of NSP2, and 305, 189, 519, or 336 in the presence of 5 nM, 10 nM, 20 nM or 200 nM NSP2, respectively; 370 in the absence of NSP2-ΔC, and 346, 355, 790, or 590 in the presence of 5 nM, 10 nM, 20 nM or 200 nM NSP2-ΔC, respectively.

5.3 Discussion and Outlook

Figure 5.5 shows our proposed model to explain why NSP2-ΔC simultaneously exhibits enhanced RNA-unwinding and reduced RNA-annealing activity. NSP2-ΔC unfolds RNA efficiently, forming a highly stable complex. As this complex is energetically favorable, the release of unfolded RNA is considerably slowed down. This, in turn, hinders binding of complementary RNA strands to promote RNA-RNA interactions and annealing. Based on smFRET experiments and additional biochemical characterizations of the NSP2-RNA interaction, we were able to show that the acidic patch DEED within the CTR is required to release RNA presumambly via charge repulsion to prevent the thermodynamically trapped state. Although the presence of the CTR reduces the helix-unwinding activity of NSP2, it is essential for its full RNA chaperone activity, including chaperone recycling.

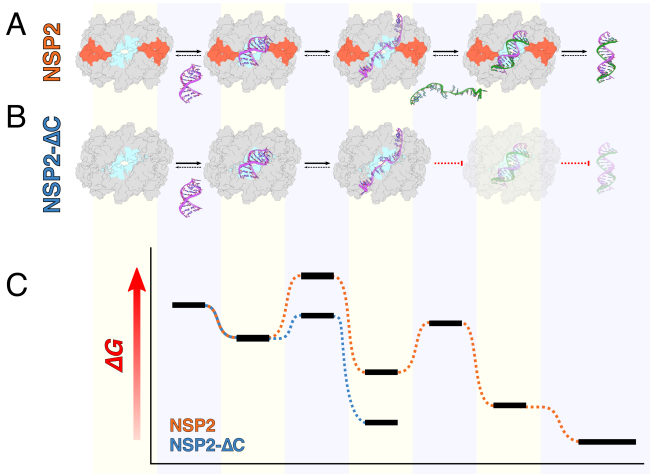


FIGURE 5.5: Proposed machanism of CTD-accelerated RNA dissociation from NSP2 **A** CTR (in orange) is required for efficient RNA-RNA interaction and release from NSP2. **B** NSP2-ΔC lacking the CTR forms a highly stable complex with RNA and hinders its release. **C** Proposed free energy diagram of NSP2 and NSP2-ΔC chaperone activity. Figure taken from Bravo et al., 2020.

Interestingly, the importance of the CTR in RNA displacement has been previously shown for another multimeric RNA chaperone protein, bacterial Hfq (Santiago-Frangos et al., 2016). It is thus possible that CTR-assisted RNA release represents a more general mechanism of RNA chaperone regulation in multimeric RNA chaperones.

Chapter 6

Single-molecule Förster Resonance Energy Transfer reveals conformational changes of bacterial adhesin SdrG upon ligand binding

For this project, smFRET experiments performed by me are combined with molecular dynamics (MD) simulations contributed by my colleague Simon Wanninger to quantitatively describe the conformational states of bacterial adhesion protein SdrG in the presence and absence of its target peptide Fgβ. The project is still ongoing and a manuscript in preparation based on the findings presented in this chapter and additional planned experiments described in the outlook section of this chapter (section 6.4).

6.1 Motivation

The opportunistic pathogen *Staphylococcus epidermis* is the major cause of nosocomial infections, typically initiated by the attachment of the microbe to the host tissue (Otto, 2009). For this attachment, *S. epidermis* requires the adhesin SD-repeat protein G (SdrG), which targets the N-terminus of the human fibrinogen β chain (Fgβ) to form a mechanostable complex that withstands high forces (Milles et al., 2018). For the development of inhibitors of *S. epidermis* infections, it is critical to understand its binding mechanism.
SdrG is composed of three subdomains termed N1, N2 and N3, the latter being found at the C-terminus of the protein. The ligand binding site of SdrG has been located to subdomains N2 and N3, while the N-terminal N1 domain is usually cleaved proteolytically *in vivo*. To investigate the binding mechanism of SdrG, X-ray crystallographic experiments were performed (Ponnuraj et al., 2003) and identified the C-terminal extension of the N3 domain, termed locking strand, as the key player in the transition from an open to a closed, peptide-bound state. A dock, lock and latch (DLL) binding mechanism was proposed: the target peptide first docks into the open binding trench between the SdrG subdomains N2 and N3 and is then fixed in place by the locking strand. In a final step, the locking strand is connected to the N2 domain through ß-strand completion, forming a latch and thus the closed conformation. While this latch is clearly absent in the open state, the exact conformation of the locking strand is still unclear, as its C-terminal amino acids could not be resolved in the crystal structure. The movement of the locking strand was later investigated by FRET (Bowden et al., 2008). These ensemble data indicated a close proximity of the locking strand to the N3 domain - from which it originates - in the absence of Fgβ, and a conformational change after peptide addition, in good agreement with the DLL mechanism. However, as only two FRET pairs and thus two intramolecular distances were probed on the ensemble level, the description of the conformational changes remained qualitative.
Here, we use a combination of smFRET and all-atom MD simulations to quantitatively describe the conformational states of SdrG in the absence and presence of Fgβ and to resolve the question whether the locking strand is dynamic or adopts a fixed position in the open state. Placing a total of 7 FRET pairs on SdrG allows us to probe the locking strand position and motion with respect to subdomains N2 and N3, and to scan for inter-domain dynamics, as illustrated in Figure 6.1.

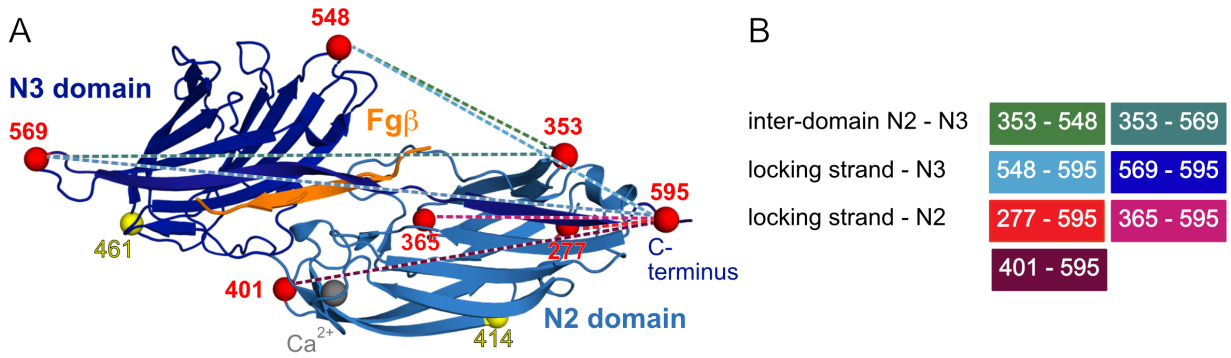


FIGURE 6.1: Structure of $SdrG_{274-596}$ and design of the SdrG FRET sensors. **A** The structure of Fgβ (orange) bound to SdrG (blue) (PDB: 1R17, Ponnuraj et al., 2003). The C-terminal locking strand of SdrG locks the peptide in the binding pocket between the subdomains N2 (light blue) and N3 (dark blue). Red spheres indicate the Cα-atoms of the residues that were mutated to cysteines for stochastic labeling with Atto532 and Atto643 or Alexa488 and Atto643. The labeling combinations investigated by smFRET are shown as dashed lines. Yellow spheres respresent Cα-atoms of residues for the additional control experiments mentioned later in this chapter (section 6.4). The grey sphere indicates a Ca^{2+}-ion. **B** Overview of the labeling combinations designed to probe for inter-domain dynamics and motions of the locking strand towards subdomains N2 and N3.

6.2 Experimental procedures

6.2.1 Gene construction

SdrG genes for the different double-cysteine mutants were constructed by our collaboration partner Lukas Milles. The *S. epidermis* SdrG N2 N3 domain genes (UniProt: Q9KI13) had been synthesized codon-optimized for expression in *E. coli* as linear DNA fragments (GeneArt, ThermoFisher Scientific, Regensburg, Germany) as described previously (Milles et al., 2018). Genes were inserted into pAC4 vectors and a hexahistidine-, ybbR- and AviTag included via Gibson assembly (New England Biolabs, Massachusetts, USA). All point mutations of amino acids to cysteines, which were later used as attachment points for the fluorophores, were created through polymerase chain reactions (Phusion Polymerase, New England Biolabs, Massachusetts, USA) with appropriate primers, followed by blunt-end ligation cloning using the T4 Ligase (Thermo Scientific, Massachusetts, USA) or Gibson assembly. The resulting open reading frames of all constructs were verified by DNA sequencing (Eurofins Genomics, Ebersberg, Germany).

6.2.2 Protein expression and purification

SdrG mutants were expressed in the competent *E. coli* strain CVB101 (Avidity Biotech) and purified on a Ni-NTA column (HisTrap FF 5mL on a Äkta Start system, both GE Healthcare, Massachusetts, USA) via their hexahistidine-tag by our collaboration partner Lukas Milles. A detailed protocol has been previously described (Milles et al., 2018).

6.2.3 Fluorescence labeling of double-cysteine SdrG mutants

For fluorescent labeling of SdrG, cysteines were introduced in the N2 domain (residues 276-424, numbering as per PDB 1R17) at positions 277, 353, 365 and 401, and in the N3 domain (residues 425-596) at positions 548, 569 and 595, with position 595 being located on the locking strand close to the C-terminus (Figure 6.1A). A total of 7 double-cysteine mutants were generated and stochastically labeled with maleimide-functionalized dyes. The cysteine combinations were chosen to either probe the distance between domains N2 and N3 or to investigate the motion of the locking strand with respect to N2 and N3 (Figure 6.1B). As recombinant SdrG does not contain any cysteine residues, the introduced cysteines are the only available attachment points for the dyes. The protein was incubated with a 2-fold molar excess of the maleimide-functionalized dyes Atto532 and Atto643 or Alexa Fluor 488 and Atto643 (Atto-Tec, Siegen, Germany or Thermo Scientific, Massachusetts, USA) for three hours at room temperature in the presence of 1 mM TCEP. Excess dye was removed by ultrafiltration using molecular weight cut-off filters of 10 kDa (Merck Millipore, Burlington, Massachusetts, USA).

6.2.4 AFM single-molecule force spectroscopy

AFM control experiments using the double-labeled SdrG mutants were performed by our collaboration partner Lukas Milles. Glass surfaces and cantilevers were cleaned and coated with heterobifunctional α-Maleinimido-hexanoic-PEG-NHS (Rapp Polymere, Tübingen, Germany) and incubated with Coenzyme A to covalently pull down SdrG proteins via their ybbR-tag. AFM single-molecule force spectrospcopy measurements were conducted at room temperature (approximately 25°C) and analyzed as described previously (Milles et al., 2018). Typically, 50.000-100.000 curves were recorded for each double-labeled SdrG mutant.

6.2.5 SmFRET measurements in solution and on the surface

SmFRET measurements were performed in solution and using surface-immobilization.
Solution measurements were performed in 8- or 4-well chamber slides (Nunc Lab-Tek, VWR), which were coated with BSA (New England Biolabs, 20 mg/ml diluted to 1 mg/ml in PBS) on one of the PIE-MFD confocal microscopes (introduced in section 3.1.2). Labeled SdrG constructs were diluted to concentrations of 50-150 pM in imaging buffer (PBS and 1 mM Trolox to reduce photobleaching, Cordes et al., 2009) and measured in the absence or presence of the unlabeled target peptide Fgβ for up to 5 hours. The concentration of Fgβ was varied between 125 nM and 12.5 μM (roughly 10^3-fold to 10^5- fold molar excess). Data were analyzed using the software *PAM* (Schrimpf et al., 2018). Bursts were selected using the all-photon burst search algorithm, which is based on a sliding time window approach (Nir et al., 2006) as described in section 3.1.6. Parameters for the burst search were set to a minimum of 100 photons per burst, a time window of 500 µs and 10 photons per time window. Incompletely labeled molecules were removed using typical thresholds < 10 for the ALEX-2CDE filter (Tomov et al., 2012). Correction factors for accurate FRET effciencies were calculated according to standardized procedures (Hellenkamp et al., 2018; Kudryavtsev et al., 2012) and are summarized in Table 6.1 and Table 6.2 for the dye combinations Atto532-Atto643 and Alexa488-Atto643, respectively. Correction factors were determined for every double-cysteine SdrG mutant individually to account for potential effects of the labeling positions on the dyes as well as for slight changes in setup alignment. Distances from the experimental data were calculated using the photon distribution analysis (PDA) (Antonik et al., 2006) (see section 3.1.6).

TABLE 6.1: Correction factors α, β, γ and δ for the smFRET burst analysis of Atto532- and Atto643-labeled SdrG FRET sensors.

SdrG label positions	crosstalk α	β-factor	γ-factor	direct excitation δ
353-548	0.0202	0.9317	0.4533	0.0585
353-569	0.0212	1.1498	0.4982	0.0644
548-595	0.0226	0.8748	0.4841	0.0666
569-595	0.0225	1.1570	0.4705	0.0577
277-595	0.0222	0.8748	0.5350	0.0633
365-595	0.0225	0.9317	0.4405	0.0649
401-595	0.0210	1.2044	0.5333	0.0497

TABLE 6.2: Correction factors α, β, γ and δ for the smFRET burst analysis of Alexa488- and Atto643-labeled SdrG FRET sensors.

SdrG label positions	crosstalk α	β-factor	γ-factor	direct excitation δ
353-548	0.0064	1.7596	0.4976	0.0363
353-569	0.0022	2.0830	0.3868	0.0128
548-595	0.0049	1.7872	0.7805	0.0345
569-595	0.0018	1.3710	0.5197	0.0520
277-595	0.0156	1.9111	0.5930	0.0513
365-595	0.0074	1.7596	0.4976	0.0363
401-595	0.0030	1.7992	0.4985	0.0226

Experiments on the objective-type TIRF setup were performed in flow cells (see Section 3.2.3). The surface of the flow chambers was passivated with PEG and a small fraction of PEG-biotin. After incubation with streptavidin (Sigma-Aldrich, 0.2 mg/ml in PBS) for 15 min, the flow chambers were washed with PBS. The protein samples, which carried a biotin-tag at the C-terminus (AviTag, Avidity LLC, biotin-labeled during

expression in Avidity CV101 strain), were diluted to 100 pM in PBS and immobilized on the coverslip through the biotin-streptavidin-biotin binding. Unbound proteins were washed away with PBS. For stabilization of the fluorophores and oxygen removal, a buffer composed of PBS, 1 mM Trolox (UV activated for 5 min), 10 % (v/v) glycerol, 1 % (w/v) glucose and 10 % (v/v) glucoseoxidase-catalase solution (Stein et al., 2012) was used. The unlabeled target peptide Fgβ was diluted in this photo-stabilizing buffer and added to the flow chamber. Data were recorded with an alternating excitation sequence of green-red with typical laser powers of 15 mW for the green 532 nm laser and 8 mW for the red 647 nm laser measured before the objective at an exposure time of 100 ms for 1000 frames. Data were analysed using the Matlab-based programs *Tracy* and *TRacer* (see section 3.2.5).

6.2.6 All-atom molecular dynamics (MD) simulations

All-atom MD simulations were performed by Simon Wanninger and are based on the crystal structure of the SdrG:Fgβ complex (PDB: 1R17, Ponnuraj et al., 2003). Simulations were performed using the open-source software GROMACS 5.1.4 (Abraham et al., 2015), the a99SB-*disp* force field (Robustelli et al., 2018) and the TIP4P-D water model (Piana et al., 2015) at a temperature of $T = 300K$. Short-range non-bonded interactions exceeding distances of 10 Å were removed using the Verlet neighbor search algorithm (Verlet, 1967). Long-range electrostatic interactions were described with the particle-mesh Ewald (PME) method (Darden et al., 1998). Simulations were performed at a time step of 2 fs with total simulation times of 200 ns for the SdrG:Fgβ complex and 1200 ns for the SdrG apoprotein. As we expected a dynamic behavior of the locking strand in the open conformation of SdrG, we chose a longer simulation time in this case.

6.2.7 AV calculations

To compare the experimental to theoretical distances extracted from the crystal structure of SdrG:Fgβ (PDB: 1R17, Ponnuraj et al., 2003) or from the MD simulations, the accessible volumes for Atto532 and Atto643 attached to SdrG were calculated with the FRET positioning and screening software (Kalinin et al., 2012) using the Cα-atoms of the respective cysteine residues as anchor points, the three radii AV3-model and the dye parameters summarized in Table 6.3. In the AV3-model, three spherical AVs with radii R_1, R_2 and R_3 are calculated seperately and the grid points in the final model weighted according to the numbers of AVs they belong to. To compute FRET-averaged distances $\langle R_{DA} \rangle_E$ and mean FRET efficiencies $\langle E \rangle$, a Förster radius of 59 Å was used, which is the theoretical value for the dye pair Atto532-Atto643 given by the manufacturer, calculated with the orientation factor $\kappa^2 = \frac{2}{3}$ and a donor quantum yield of $\phi_D = 0.9$ (lifetime $\tau_D = 3.8$ ns).[1] We chose this theoretical value, as donor lifetimes in our experiments were close to the theoretical value (we thus assume no influence on the photophysical properties of Atto532 in our system) and as fluorescence anisotropy measurements indicated no sticking of the dyes to SdrG.

TABLE 6.3: Dye parameters for accessible volume calculations with the AV3-model.

dye	linker length [Å]	linker width [Å]	R_1 [Å]	R_2 [Å]	R_3 [Å]
Atto532-maleimide	20.5	4.5	5.5	4.5	1.5
Atto643-maleimide	21	4.5	7.15	4.5	1.5

6.2.8 Structural modeling

To identify those structures in our all-atom MD simulations that are in good agreement with the measured smFRET data, the weighted data-model deviation χ^2 (as defined in Kalinin et al., 2012) was calculated

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(R_{DA(i)} - R_{model(i)}\right)^2}{\Delta R_{DA(i)}^2} \tag{6.1}$$

where N is the number of measured distances (i.e. the degrees of freedom), R_{DA} the distance calculated from the FRET measurement, R_{model} the distance obtained from the MD simulation and ΔR_{DA} the uncertainty of the measured data. ΔR_{DA} depends on systematic errors during the data analysis, the uncertainty of using the correct orientation factor κ^2, as well as the uncertainty of using the correct dye model for AV calculations

[1] information taken from: https://www.atto-tec.com (accessed September 20, 2019)

(Dimura et al., 2016). Here, we used $\Delta R_{DA} = 3$ Å for all measurements except for the high-FRET efficiency state measured for construct 548-595, for which we used $\Delta R_{DA} = 5$ Å.

The reduced χ^2 value was calculated by deviding by N

$$\chi^2_{red.} = \frac{1}{N}\chi^2 \tag{6.2}$$

6.3 Results

6.3.1 AFM investigation of the SdrG:Fgβ complex

To investigate a potential influence of the fluorescent labels on the binding of SdrG to Fgβ, AFM control experiments were performed by Lukas Milles comparing wild-type (wt) and fluorescently labeled double-cysteine SdrG (Figure 6.2A). The single-molecule force spectroscopy measurements confirmed that rupture forces for Sdrg:Fgβ complexes are almost identical for wt and labeled SdrG (green and grey histograms in Figure 6.2B, respectively). These findigs confirm that fluorescently-labeled SdrG is not only able to bind to Fgβ, but that the complexes are as mechanostable as in the absence of the dyes. However, in our experiments, not all molecules are 100 % labeled with both donor and acceptor fluorophores, as labeling efficiencies for the single dyes were typically in the range of $30\% - 60\%$. While SdrG molecules that carry no, a single or both dyes could not be distinguished in the AFM experiment, the absence of additional populations in the rupture force histograms indicates a similar behavior for the different molecules and thus no obvious influence of the dyes. However, a modulation of the binding affinity could not be measured by AFM.

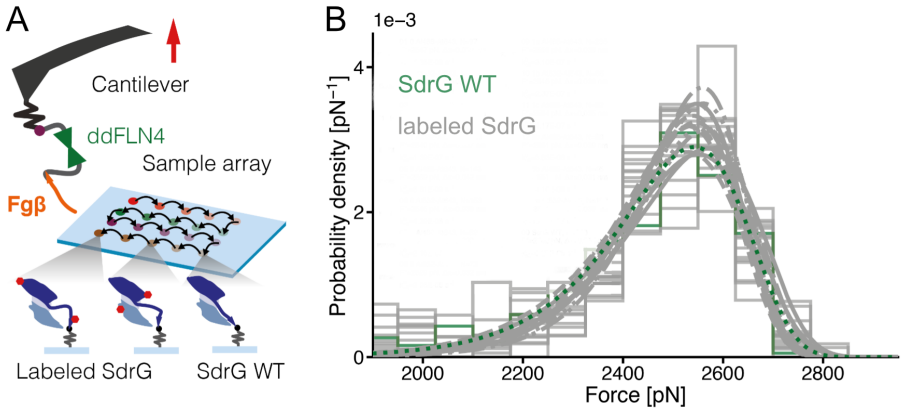


FIGURE 6.2: Fluorescently-labeled SdrG double-cysteine mutants form mechanostable complexes with the target peptide Fgβ comparable to those with wt SdrG. **A** A schematic of the experimental AFM setup including the ddFLN4 fingerprint domain (in green). Wt SdrG and fluorescently-labeled double-cysteine SdrG mutants are covalently bound to the glass surface via PEG linkers and their ybbR-tag. Both SdrG and Fgβ are force-loaded from their respective C-termini. The AFM cantilever is retracted at constant velocity until the Sdrg:Fgβ complex breaks. The glass surface is then moved to a new spot and the measurement repeated. **B** Rupture forces for wt and labeled SdrG binding to Fgβ (green and grey histograms, respectively). Shown are data acquired using SdrG double-cysteine mutants labeled with Alexa488 and Atto643, or with Atto532 and Atto643. No significant difference in rupture force and thus complex stability is detected for the various measurements. Figure provided by Lukas Milles.

6.3.2 smFRET investigation of the SdrG:Fgβ complex

Having confirmed the biological functionality of the double-labeled SdrG mutants, smFRET measurements were performed next to test how well we can reproduce the distances derived from the closed conformation of the SdrG:Fgβ crystal structure (Ponnuraj et al., 2003) in our experiments. SmFRET experiments were first performed in solution using the dye combination Atto532-Atto643 (Figure 6.3A). To realize the closed, peptide-bound state, SdrG was measured in the presence of a high molar excess of Fgβ (12.5 μM, $\sim 10^5$-fold molar excess), well above the dissociation constant (K_d) of $\sim$ 400 nM (Ponnuraj et al., 2003). The resulting molecule-wise FRET efficiency histograms in Figure 6.3B show a single population for most mutants and thus indicate a well-defined protein conformation in the presence of Fgβ. We compared intramolecular distances extracted

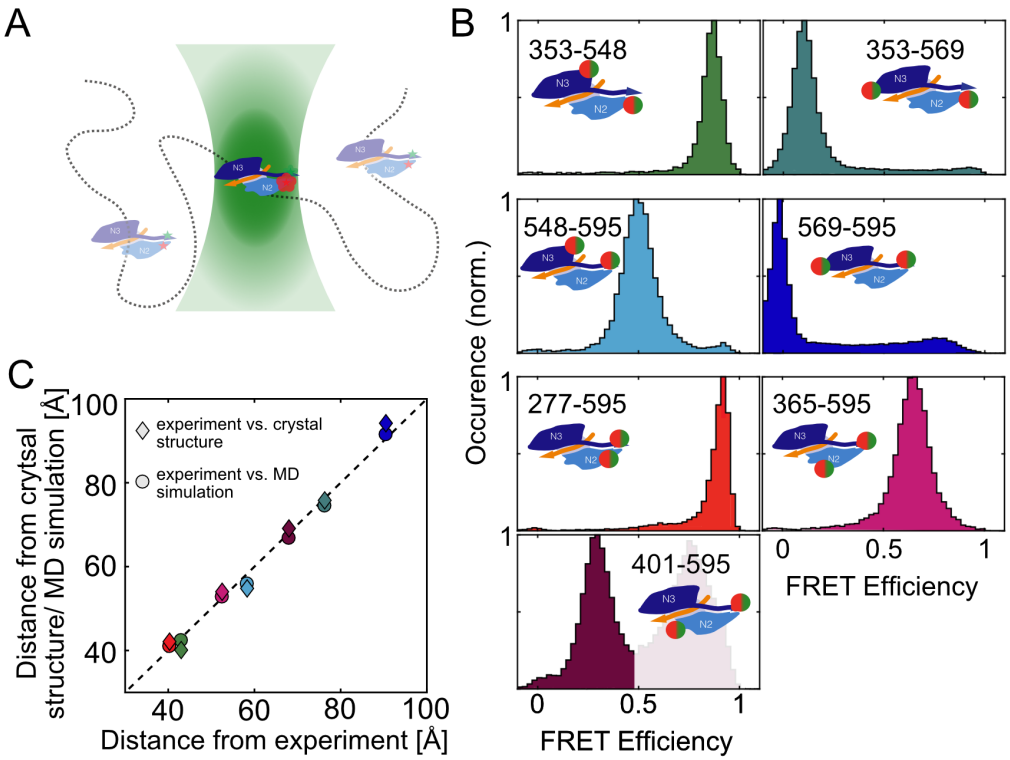


FIGURE 6.3: Conformation of the SdrG:Fgβ complex studied via smFRET in solution. **A** A schematic of a double-labeled SdrG:Fgβ complex diffusing through the confocal volume. The energy transfer efficiency between the donor dye Atto532 (green star) and the acceptor dye Atto643 (red star) is calculated for every single-molecule event. **B** Histograms of the molecule-wise smFRET efficiency for the seven investigated SdrG mutants, shown as cartoons. The locations of the cysteine residues used for stochastic labeling are indicated by green-red circles. **C** A comparison of the measured, simulated and theoretical distances for the different constructs extracted from the smFRET experiments by PDA, from all-atom MD simulations of the SdrG:Fgβ complex by AV calculations and from the SdrG:Fgβ crystal structure (PDB 1R17) by AV calculations, respectively. A Förster radius of 59 Å was used to convert experimental FRET efficiencies into distances. A linear correlation is observed between the experimental and theoretical values (indicated by circles) and between the experimental and simulated values (indicated by diamonds). The color code corresponds to the smFRET efficiency histograms in B.

from our smFRET data with the SdrG:Fgβ crystal structure (PDB 1R17). To obtain distances from the crystal structure, the sterically accessible volume (AV) of the FRET dyes was determined for each labeling position and the average FRET efficiency values were calculated (Kalinin et al., 2012). Distances from the experimental data were calculated using the photon distribution analysis (PDA) (Antonik et al., 2006). For most constructs, a single peak and thus dominant conformation is observed. In the case of construct 401-595 (Figure 6.3B, burgundy color) a second pronounced population is present, which we interpreted as a mixture of free and Fgβ-bound SdrG. Indeed, by comparing the measured to the theoretical values, we could identify the lower FRET efficiency population at $E = 0.3$ as the peptide-bound conformation, while the population at $E = 0.7$ was later confirmed to represent the conformation of the apoprotein (Figure 6.4).
Overall, the smFRET results of the peptide-bound conformation are in good agreement with the theoretical values with a root-mean-square deviation (RSMD) of 1.3 Å (Figure 6.3C), which lies within the expected error range for smFRET experiments (Hellenkamp et al., 2018). As an additional control, we performed all-atom MD simulations of the SdrG:Fgβ complex and extracted distances by AV calculations. The values from the simulation match the experimentally determined distances with a RSMD of 2.4 Å (Figure 6.3C), and are also in good agreement with the theoretical data from the crystal structure (RSMD of 1.9 Å). Thus, the distances derived from the closed conformation of the SdrG:Fgβ crystal structure could be reproduced through our smFRET experiments and the all-atom MD simulation.

6.3.3 SmFRET investigation of the SdrG apoprotein

Having validated our experimenatl approach, we next investigated the conformation of SdrG as an apoprotein. To this end, smFRET experiments were performed in the absence of Fgβ. The resulting FRET efficiency

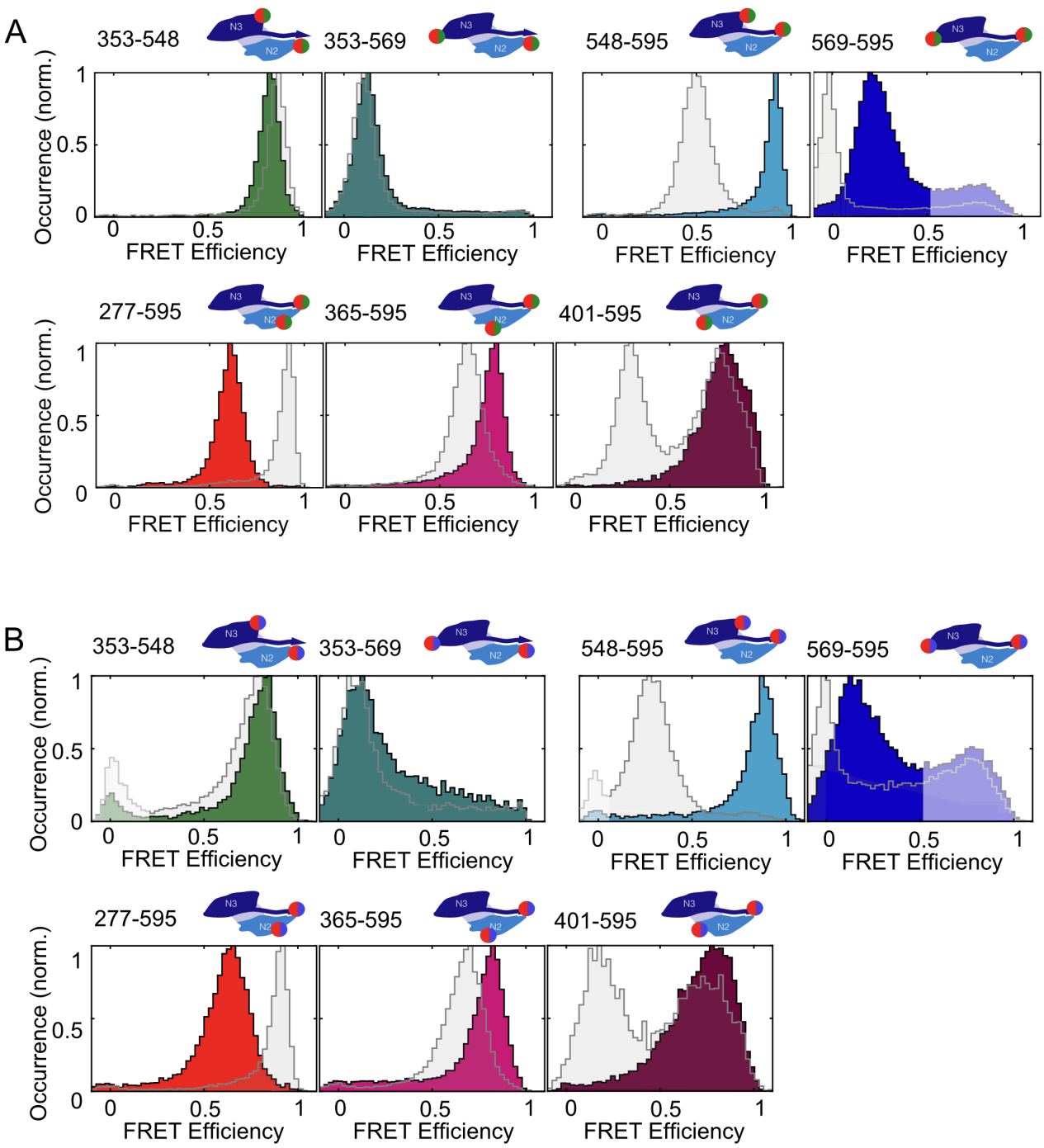


FIGURE 6.4: Conformation of the SdrG apoprotein studied by smFRET in solution. SmFRET efficiency histograms obtained in the absence of Fgβ (in color) are compared to smFRET efficiency histograms of the SdrG:Fgß complex (in grey). Significant changes in FRET efficiency are detected for mutants with a label positioned on the locking strand at residue 595 and a label in the N2 or N3 domain (histograms in shades of red and blue, respectively). Experiments were performed using the dye combination Atto532-Atto643 (histograms in **A**) and Alexa488-Atto643 (histograms in **B**). The low-FRET efficiency populations for constructs 353-548 (green) and 548-595 (blue) in **B** indicate donor blinking, which is more pronounced in the presence of Fgβ.

histograms for the dye combination Atto532-Atto643 in Figure 6.4A display one main population for each construct (in color) and are compared to the FRET efficiency histograms of SdrG in complex with Fgβ from Figure 6.3C (here shown in grey). The inter-domain sensors with one label positioned in the N2 and one in the N3 domain (residues 353-548, in green, and residues 353-569, in teal) show almost identical FRET efficiency histograms and thus similar intramolecular distances in the presence and absence of Fgβ. In contrast, a clear shift

in FRET efficiency was observed for all constructs with a label on the locking strand (residue 595). This shift was present for mutants with the second label in the N3 domain (residues 548 and 569; histograms in shades of blue) as well as for mutants with the second label in the N2 domain (residues 277, 365 and 401; histograms in shades of red). To confirm that our observations are not dependent on the dye pair used (Atto532-Atto643), we repeated the measurements using the dye pair Alexa488-Atto643 on SdrG and were able to observe the same trends, as shown in Figure 6.4B. Changes in FRET efficiency are thus associated with the motion of the locking strand.

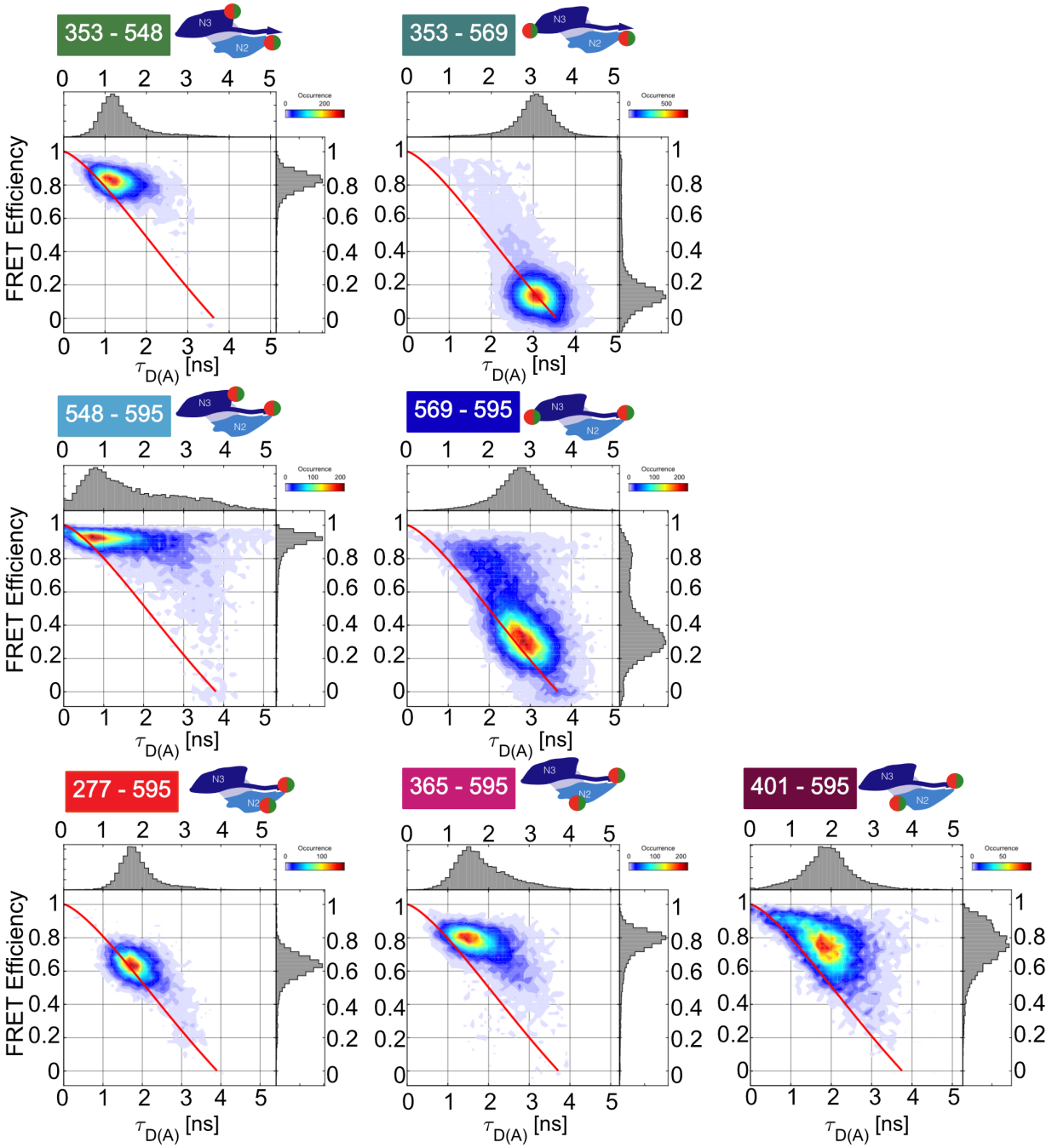


FIGURE 6.5: Conformational dynamics of the locking strand indicated in the absence of Fgβ. Two-dimensional histograms of the FRET efficiency versus the donor lifetime show a clear deviation from the static-FRET line (shown as a solid red line) for SdrG mutants with a label on the locking strand (residue 595). Data are based on SdrG labeled with Atto532 and Atto643.

As mentioned above, it was not possible for the 401-595 construct to completely shift the conformation to the peptide-bound, closed state in the presence of a high molar excess of Fgβ, which results in the two FRET efficiency populations observed in Figure 6.3C (burgundy histogram; also shown in Figure 6.4 in grey). One possible explanation is that the fluorescent label at position 401 affects the binding affinity. Another labeling position that seems to interfere with the structural integrity of SdrG is residue 569, which is located on a flexible loop within the N3 domain (Figure 6.3A). For construct 353-569, a bridging towards high-FRET efficiencies is observed in the absence of Fgβ (teal histograms in Figure 6.4). To realize an inter-dye distance small enough to explain the high-FRET efficiency peak (~ 30 Å), the two domains N2 and N3 would have to collapse. As the high-FRET efficiency values persist in the presence of Fgβ, dysfunctional structures are indicated. For construct 569-595, two FRET efficiency populations are observed in the absence of Fgβ (cobalt blue histogram in Figure 6.4). The high-FRET efficiency population is slightly reduced in the presence of Fgβ, but does not disappear completely. As its occurrence varied for different protein expression batches and as none of the other constructs indicated two different conformations of the apoprotein, we attributed the high-FRET efficiency state to a labeling artifact.

In accordance with crystallographic data (Ponnuraj et al., 2003) and FRET experiments on the ensemble level (Bowden et al., 2008), our smFRET data confirm different conformational states of the locking strand in the presence and absence of Fgβ, while the conformation of the two subdomains does not change significantly. Although the mobility of the locking strand is a prerequisite for the DLL mechanism and obviously defines the transition from an open to a closed state, we did not observe an equilibrium of the open and closed conformations for the SdrG apoprotein, as previously hypothesized (Bowden et al., 2008). This also rules out that the capture of the target Fgβ peptide is aided by stochastic opening and closing of the locking strand over the binding cleft. However, dynamics on the sub-millisecond time scale are indicated for the open conformation. Histograms of FRET efficiency versus donor lifetime $\tau_{D(A)}$ in Figure 6.5 show a rightward shift from the static-FRET line for SdrG constructs with a label on the locking strand (second and third rows). Interestingly, the shift from the static-FRET line is less significant for construct 277-595, indicating that the label at position 277 is less sensitive in detecting dynamic fluctuations of the locking strand. As expected, no deviation from the static-FRET line is observed for the inter-domain sensor 353-569, as these label positions are insensitive to structural changes of the locking strand. The slight shift from the static-FRET line for the second inter-domain sensor 353-548 is most likely due to the close proximity of the FRET dyes in this construct, as direct contact between the dyes can cause acceptor blinking and formation of quenched states. Thus, our data suggest that the locking strand is flexible in the absence of Fgβ, while adopting a conformation that clearly differs from its closed state in the SdrG:Fgβ complex.

6.3.4 Proposed structural model for the open conformation of SdrG

To find a structural model for the open conformation, we compared FRET-derived distances to a MD simulation of the SdrG apoprotein. Although a crystal structure of SdrG in the absence of Fgβ exists (PDB 1R19, Ponnuraj et al., 2003), we could not extract distances for comparison with our smFRET data, as almost all residues chosen as labeling positions are located in unresolved parts of the protein. However, having shown a good agreement of our all-atom MD simulation with the crystal structure of SdrG:Fgβ, this inspired confidence to model the structure of the apoprotein based on our FRET data. To this end, we simulated an ensemble of structures in the open conformation. As a starting point, the crystal structure of SdrG:Fgβ (PDB 1R17) was chosen, the peptide deleted and the C-terminus of the locking strand (residues 581-596) pulled away from its binding trench. For every frame of the simulation, intramolecular distances were extracted and compared to the FRET data. The structures that fit the experimental values best ($\chi^2_{red.} < 1.5$, Figure 6.6A), are presented in Figure 6.6B. As the locking strand undergoes fast dynamic conformations in the absence of Fgβ, these structures represent kinetically averaged positions. The locking strand, which is colored in red for better visibility, points away from the two domains. Its C-terminal part is unbound and shows a similar orientation for most selected structures. The absence of a specific binding site is supported by the fact that it has not been possible to resolve the C-terminal part of the locking strand in the crystal structure (Ponnuraj et al., 2003). The inset of Figure 6.6B shows the previous model of the SdrG apoprotein with locking strand residues 581-596 missing. Thus, our model does not only confirm the extension of the locking strand in solution, but also indicates the preferred position of its C-terminus in the open conformation of SdrG and proves the absence of specific interactions.

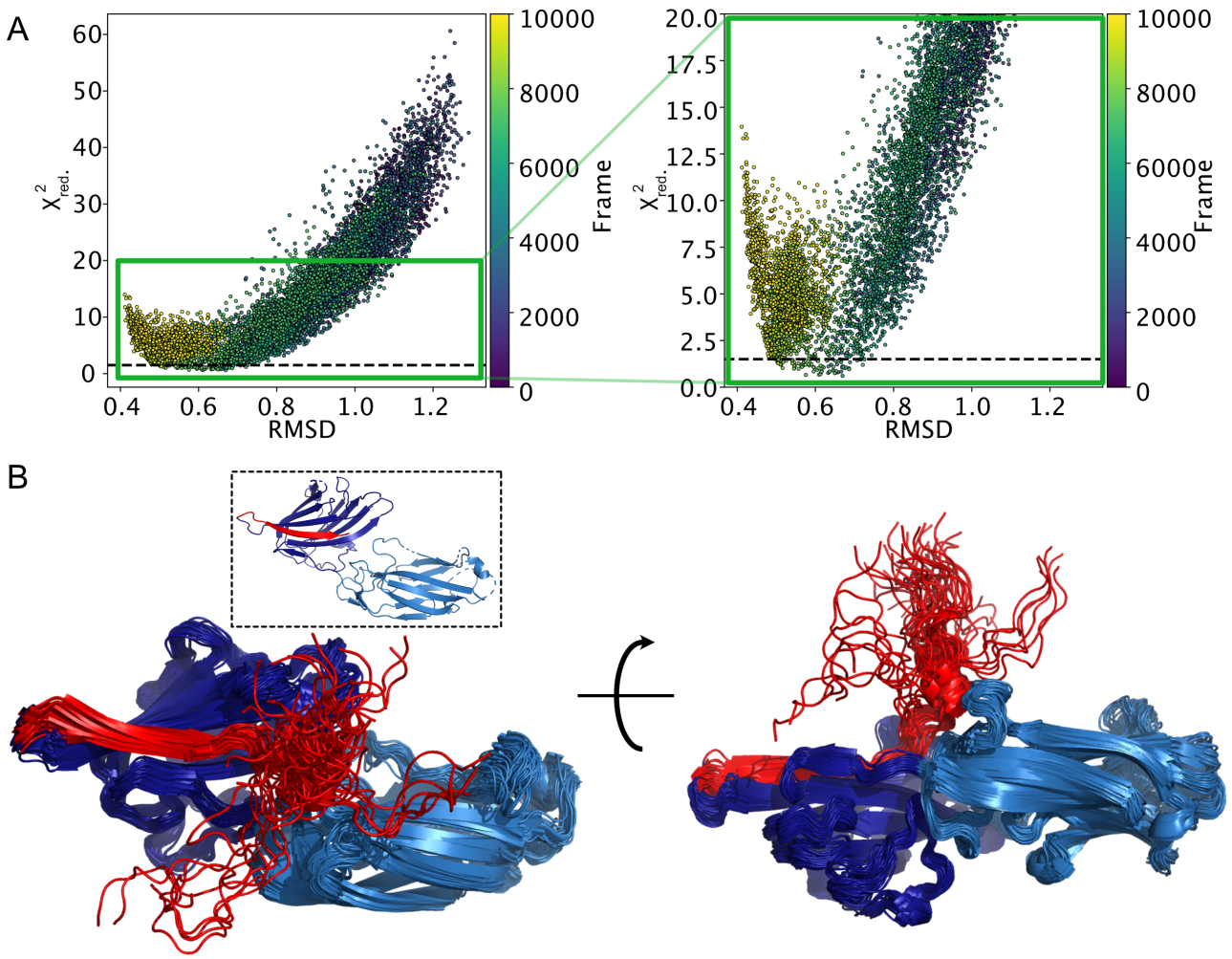


FIGURE 6.6: Structural model of the SdrG apoprotein based on smFRET data and all-atom MD simulations. **A** Structural models resulting in $\chi^2_{red.} < 1.5$ (indicated by the dashed black line) were selected to model the SdrG apoprotein. Plotted on the x-axis is the RMSD comparing the conformation of the SdrG apoprotein extracted from the MD simultations and the conformation of the SdrG:Fgβ complex based on the crystal structure (PDB 1R17, Ponnuraj et al., 2003). **B** Shown are two perspectives of 57 overlayed structures that fit the experimentally determined inter-dye distances ($\chi^2_{red.} < 1.5$). The subdomains N2 and N3 are colored in light and dark blue, respectively, and the locking strand (residues 569-596) is shown in red. Screening was performed by Simon Wanninger. The inset shows the previous structural model based on X-ray crystallography, which could not resolve the C-terminal part of the locking strand (PDB 1R19, Ponnuraj et al., 2003).

6.3.5 Switching between an open and closed conformation of SdrG occurs on the second timescale in the presence of Fgβ

While no interconversion between the open and closed conformations was indicated in the absence of Fgβ, we observed a co-existence of two FRET efficiency populations at low concentrations of Fgβ (125 nM and 1.25μM). At higher peptide concentrations, the equilibrium shifted towards the closed state for all investigated SdrG FRET sensors (Figure 6.7). The shifts are most distinct for constructs with a label positioned on the locking strand at residue 595 (histograms in shades of blue and red) with black and orange dashed lines highlighting the FRET efficiencies associated with the open and closed states, respectively.

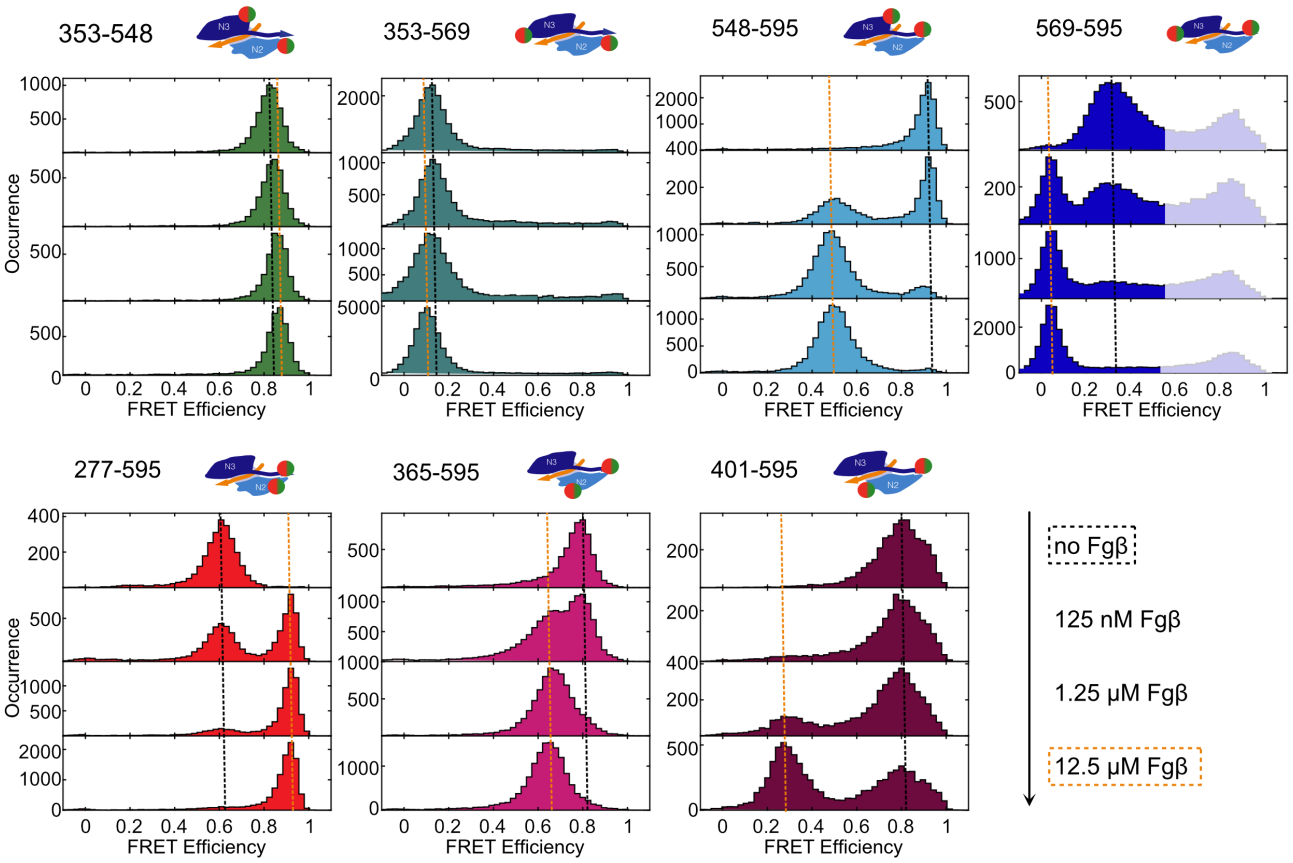


FIGURE 6.7: SmFRET efficiency histograms obtained in the absence and at different concentrations of Fgβ for the dye combination Atto532-Atto643. Increasing concentrations of Fgβ (from top to bottom) shift the equilibrium from the open conformation of SdrG (black dashed line) to the closed conformation (orange dashed line).

No or only little dynamic conversion of the two states was indicated on the time scale of diffusion, as exemplarily shown for SdrG constructs 548-595 and 277-595 (Figure 6.8A and B). The FRET efficiency versus donor lifetime histograms clearly illustrate donor blinking, as the deviation from the static-FRET line follows a dynamic-FRET line connecting high- and no-FRET efficiencies (indicated by the blue dashed lines). Dynamic transitions between the FRET efficiency populations of the open and closed state would follow the red dashed dynamic-FRET lines, which is not observed here.
To investigate whether dynamic switching occurs on slower time scales, smTIRF experiments were performed with SdrG mutants 277-595 and 548-595. The immobilization to a biotinylated glass surface with streptavidin using the biotin-tag at the C-terminus of SdrG allowed the observation of individual FRET sensors over a time range of seconds to minutes. The presence and photostability of both dyes was confirmed by ALEX (Kapanidis et al., 2005). The constructs were first measured in the absence of the peptide. As expected, a defined FRET efficiency population representing the open state was observed. Addition of Fgβ to surface-immobilized SdrG molecules resulted in the detection of a second FRET efficiency population, which represents the peptide-bound state. Increasing the concentration of Fgβ resulted in an increase of the closed SdrG conformation. The respective framewise histograms are shown in Figure 6.8C and D. The fact that most molecules were observed in the closed state at 10 μM Fgβ indicates that the immobilized molecules are still capable of binding the ligand (i.e. are still functional). Although the data quality and statistics for construct 277-595 were considerably better compared to experiments with construct 548-595, both SdrG mutants show similar trends. While most molecules were either detected in the open or the closed conformation in the presence of Fgβ, we observed several molecules ($\sim$ 15 %) which dynamically switched between the two states. Exemplary fluorescence intensity time traces for constructs 548-595 and 277-595 acquired in the presence of 1 μM Fgβ are shown in Figure 6.8E and F, respectively. In both cases, anti-correlated donor (GG) and FRET (GR) signals are observed, while the acceptor signal after acceptor excitation (RR) remains static. The FRET efficiency fluctuates on the timescale of seconds between two states associated with the open and closed conformation of SdrG, while no intermediate states are observed.

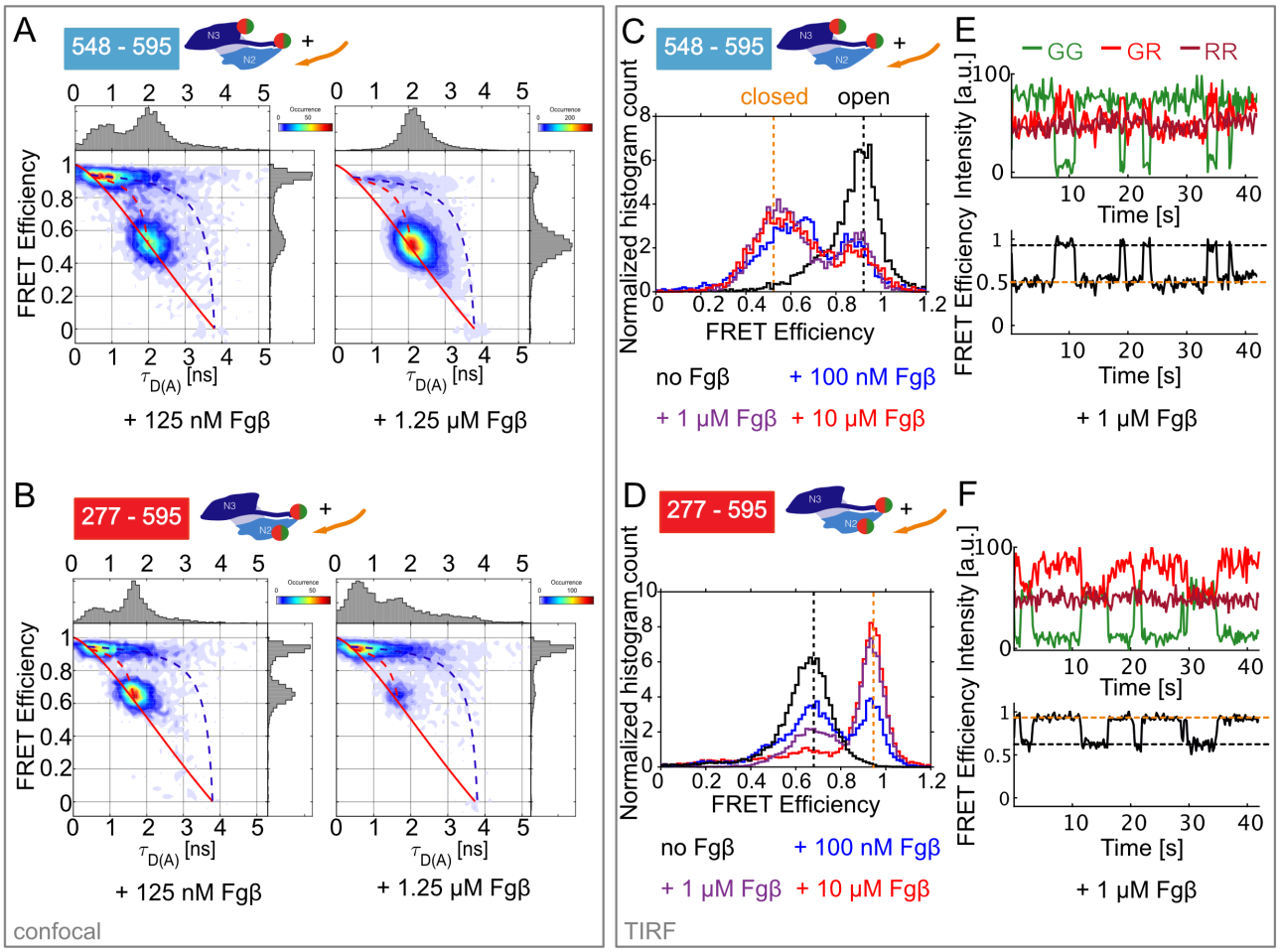


FIGURE 6.8: SdrG requires the presence of Fgβ to switch between its open and closed conformation. The slow conformational dynamics are not resolvable for freely diffusing proteins on a confocal microscope (A-B), but can be observed for surface-immobilized constructs on a TIRF microscope (C-F). **A-B** Two-dimensional histograms of the FRET efficiency versus the donor lifetime show a deviation from the static-FRET line (shown as a solid red line), which is caused by donor blinking (indicated by dashed blue lines). Conformational dynamics would follow the dashed red lines. **C-D** Framewise histograms over all recorded FRET traces show an increase of molecules in the closed conformation (indicated by an orange dashed line) with increasing concentrations of Fgβ. The numbers of analyzed molecules were 49, 46, 31 and 27 for SdrG construct 548-595 in the absence of Fgβ and in the presence of 100 nM, 1 μM and 10 μM Fgβ, respectively, and 87, 60, 73 and 63 for SdrG construct 277-595 in the absence of Fgβ and in the presence of 100 nM, 1 μM and 10 μM Fgβ, respectively. **E-F** Exemplary fluorescence intensity time traces show the conformational switching between the open and closed conformation of SdrG in the presence of 1 μM Fgβ. The green signal after green excitation (GG) is shown in green, the red signal after green excitation (GR, i.e. FRET signal) in red, and the red signal after red excitation (RR) in dark red. All data are based on SdrG labeled with Atto532 and Atto643.

As the switching is much slower than the diffusion of the protein in solution ($\sim$ 3 ms), these dynamic transitions could not be resolved in confocal measurements.

6.4 Discussion and Outlook

In conclusion, the combination of smFRET experiments and all-atom MD simulations allowed us to propose a structural model for the dynamic SdrG apoprotein in the absence of its target peptide Fgβ. We show that the previously unresolved C-terminus of the locking strand extends into solution and has no defined binding site. The absence of specific interactions suggests that the locking strand might act as a "fishing rod" to capture its target peptide reeling it towards its binding site. Although intrinsically dynamic on the sub-millisecond timescale, the locking strand prefers a specific orientation and does not spontaneously switch between its

open and closed states. By adding this missing piece to the structural information, we could not only confirm the central role of the locking strand (Ponnuraj et al., 2003), but could also give insights into its conformational dynamics. In contrast to an earlier hypothesis (Bowden et al., 2008), our data reveal that the presence of Fgβ is a prerequisite for the transition from the open to the closed SdrG conformation and a dynamic conversion was observed on the second timescale at sufficiently low peptide concentrations. As the locking strands only closes in the presence of Fgβ, the binding cleft is kept consistently open.

To further strengthen and refine our structural model of the SdrG apoprotein, additional double-cysteine mutants have been designed. Placing both FRET dyes into subdomain N3 (at positions 461 and 548) or into subdomain N2 (at positions 353 and 414, see Figure 6.1A for details) will give insights into potential intra-domain dynamics. We further have to clarify whether the labeling positions 401 and 569 interfere with the structural integrity of SdrG. Thus, similar FRET sensors are required to probe the movement of the locking strand with respect to the two subdomains. To this end, label combinations 461-595 and 414-595 will complement the information on the dynamic behavior of the locking strand.

In its native environment, bacterial SdrG has to adhere tightly to human Fgβ to withstand high mechanical forces and ensure a stable attachment of the microbe to the host tissue (Otto, 2009). The mechanical stress itself may significantly contribute to the formation of the tight bond. By combining smFRET with mechanical manipulation techniques such as AFM (He et al., 2012) - a method that has previously been used to characterize the extraordinarily mechanostability of the SdrG:Fgβ complex (Milles et al., 2018) - or optical tweezers (Comstock et al., 2015), conformational changes can be investigated under force to get further insights into the critical process of bacterial adhesion.

Chapter 7

Summary and Conclusion

In this thesis, FRET has been used as the main method to study the enzymatic function of three proteins on the single-molecule level.
In the first project, I used the T4 DNA ligase as a model enzyme to study the repair of DNA double-strand breaks. DNA ligases catalyze the formation of a phosphodiester bond between juxtaposed 3'-hydroxyl and 5'-phosphate ends of nucleic acids and are thus essential guardians of genomic integrity (Lehman, 1974). To bring two fluorescently-labeled DNA double strands close together in a defined geometry, I developed a DNA origami platform and examined the ligation process for freely diffusing DNA origami structures on a confocal microscope and for surface-immobilized platforms on a TIRF setup. The ligation efficiency was shown to depend on the overhang length between the two DNA double strands, their sequence and the incubation conditions. Sticky ends were in general more efficiently ligated than blunt ends. Interestingly, while complementary overhang lengths of 4 and 8 nt behaved similarly and led to comparable fractions of sealed DNA strands, 6-nt overhangs were less efficiently ligated. This suggests a preference of the T4 DNA ligase for DNA overhangs with multiples of 4 nt to be verified in future studies. I further showed that the ligation process can be followed in real time on a TIRF microscope and that the same DNA double strands can be repeatedly ligated and cut. This makes the DNA origami nanostructure a valuable platform to study a wide range of low-affinity interactions and multi-step-reactions.
The second project focused on the rotavirus protein NSP2, which is essential for genome replication and virus assembly. As a major component of viroplasms, NSP2 exhibits RNA chaperone activities ranging from RNA binding and unwinding to strand annealing and release (Bravo et al., 2018; Taraporewala et al., 2006). To better understand the underlying catalytic mechanism, C-terminally truncated and full-length mutants were investigated in this work. Similar to the DNA ligation project, I indirectly probed the enzymatic activity of the protein by using a fluorescently-labeled interaction partner - in this case a dual-labeled RNA stem-loop - and no label on NSP2 itself. The presence of NSP2 led to a clear shift from high-FRET to low-FRET efficiency in the RNA probe and thus allowed a straightforward assessment of the RNA folding state. Surprisingly, C-terminally truncated NSP2 was more efficient in RNA unwinding than the full-length protein, while its RNA annealing activity was reduced. Supported by findings from cryo-EM, structural proteomics and biochemical assays, we proposed a central role of the highly charged C-terminal region for the chaperone activity of NSP2. Although reducing the RNA unwinding activity, the presence of the C-terminal region ensures efficient RNA release via charge repulsion and thus a full cycle of chaperone activity.
In a third project, the adhesion protein SdrG from *S. epidermis* was examined, which initiates bacterial infections through its stable attachment to the host tissue (Otto, 2009). To form a complex with its target peptide Fgβ, SdrG switches from an open to a closed conformation. By placing a network of FRET pairs on SdrG, I could directly investigate its conformational states and detect dynamics in the absence and presence of Fgβ. The combination with all atom-MD simulation allowed us to propose a structural model for the dynamic SdrG apoprotein including all parts of the flexible locking strand, providing information previously inaccesible to other methods. The simultaneous assessment of structure and dynamics thus rivals established structural biology techniques like X-ray crystallography, cryo-electron microscopy or NMR spectroscopy. As these methods require the fixation of biomolecules to characterize their three-dimensional structures, information on dynamics is not available. As shown in previous studies (Dimura et al., 2016), the combination of quantitative smFRET and all-atom MD simulations is a valuable hybrid-approach to resolve the structure of dynamic biological systems and elucidate their function.
In conclusion, this thesis highlights the multiple capabilities of smFRET. The integrative approaches presented here combined this powerful tool with established structural biology techniques and computer simulations to directly connect conformational dynamics with enymatic function, while the combination with DNA nanotechnolgy allowed the study of weak and multi-step interactions that would otherwise be difficult to realize on the single-molecule level.

Bibliography

Abdisalaam, S., Davis, A. J., Chen, D. J., and Alexandrakis, G. (2014). "Scanning fluorescence correlation spectroscopy techniques to quantify the kinetics of DNA double strand break repair proteins after γ-irradiation and bleomycin treatment." *Nucleic acids research* 42.1, e5–e5.

Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". *SoftwareX* 1-2, pp. 19–25.

Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2008). "An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments". *Biophysical journal* 94.5, pp. 1826–1835.

Algar, W. R., Hildebrandt, N., Vogel, S. S., and Medintz, I. L. (2019). "FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls". *Nature methods* 16.9, pp. 815–829.

Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C. A. M. (2006). "Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis." *The journal of physical chemistry. B* 110.13, pp. 6970–6978.

Aparicio, T., Baer, R., and Gautier, J. (2014). "DNA double-strand break repair pathway choice and cancer." *DNA repair* 19, pp. 169–175.

Barth, A., Hendrix, J., Fried, D., Barak, Y., Bayer, E. A., and Lamb, D. C. (2018). "Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of Clostridium thermocellum." *Proceedings of the National Academy of Sciences of the United States of America* 13, p. 201809283.

Bartnik, K., Barth, A., Pilo-Pais, M., Crevenna, A. H., Liedl, T., and Lamb, D. C. (2020). "A DNA Origami Platform for Single-Pair Förster Resonance Energy Transfer Investigation of DNA-DNA Interactions and Ligation." *Journal of the American Chemical Society* 142.2, pp. 815–825.

Baum, L. E. and Petrie, T. (1966). "Statistical inference for probabilistic functions of finite state Markov chains". *Annals of Mathematical Statistics* 37, pp. 1554–1563.

Baum, L. E., Petrie, T., and Soules, G. (1970). "Accept Terms and Conditions on JSTOR". *Annals of Mathematical Statistics* 41, pp. 164–171.

Bernstein, D. I. (2009). "Rotavirus overview." *The Pediatric infectious disease journal* 28.3 Suppl, S50–3.

Borodavka, A., Desselberger, U., and Patton, J. T. (2018). "Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes." *Current opinion in virology* 33, pp. 106–112.

Borodavka, A., Dykeman, E. C., Schrimpf, W., and Lamb, D. C. (2017). "Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments." *eLife* 6.

Bowden, M. G., Heuck, A. P., Ponnuraj, K., Kolosova, E., Choe, D., Gurusiddappa, S., Narayana, S. V. L., Johnson, A. E., and Höök, M. (2008). "Evidence for the "dock, lock, and latch" ligand binding mechanism of the staphylococcal microbial surface component recognizing adhesive matrix molecules (MSCRAMM) SdrG." *Journal of Biological Chemistry* 283.1, pp. 638–647.

Bravo, J. P. K., Borodavka, A., Barth, A., Calabrese, A. N., Mojzes, P., Cockburn, J. J. B., Lamb, D. C., and Tuma, R. (2018). "Stability of local secondary structure determines selectivity of viral RNA chaperones." *Nucleic acids research*.

Britton, S., Coates, J., and Jackson, S. P. (2013). "A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair." *The Journal of Cell Biology* 202.3, pp. 579–595.

Comstock, M. J., Whitley, K. D., Jia, H., Sokoloski, J., Lohman, T. M., Ha, T., and Chemla, Y. R. (2015). "Protein structure. Direct observation of structure-function relationship in a nucleic acid-processing enzyme." *Science* 348.6232, pp. 352–354.

Condon, E. U. (1926). "A Theory of Intensity Distribution in Band Systems". *Physical Review* 28.6, pp. 1182–1201.

— (1928). "Nuclear Motions Associated with Electron Transitions in Diatomic Molecules". *Physical Review* 32.6, p. 858.

Cordes, T., Vogelsang, J., and Tinnefeld, P. (2009). "On the mechanism of Trolox as antiblinking and antibleaching reagent." *Journal of the American Chemical Society* 131.14, pp. 5018–5019.

Criglar, J. M., Hu, L., Crawford, S. E., Hyser, J. M., Broughman, J. R., Prasad, B. V. V., and Estes, M. K. (2014). "A novel form of rotavirus NSP2 and phosphorylation-dependent NSP2-NSP5 interactions are associated with viroplasm assembly." *Journal of virology* 88.2, pp. 786–798.

Darden, T., York, D., and Pedersen, L. (1998). "Particle mesh Ewald: An N log(N) method for Ewald sums in large systems". *The Journal of Chemical Physics* 98.12, pp. 10089–10092.

Davies, M. J. (2004). "Reactive species formed on proteins exposed to singlet oxygen." *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* 3.1, pp. 17–25.

Desselberger, U. (2014). "Rotaviruses." *Virus research* 190, pp. 75–96.

Dimura, M., Peulen, T. O., Hanke, C. A., Prakash, A., Gohlke, H., and Seidel, C. A. (2016). "Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems." *Current opinion in structural biology* 40, pp. 163–185.

Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F., and Shih, W. M. (2009). "Self-assembly of DNA into nanoscale three-dimensional shapes". *Nature* 459.7245, pp. 414–418.

Eggeling, C., Fries, J. R., Brand, L., Günther, R., and Seidel, C. A. M. (1998). "Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy". *Proceedings of the National Academy of Sciences* 95.4, pp. 1556–1561.

Engelhardt, F. A. S., Praetorius, F., Wachauf, C. H., Brüggenthies, G., Kohler, F., Kick, B., Kadletz, K. L., Pham, P. N., Behler, K. L., Gerling, T., and Dietz, H. (2019). "Custom-Size, Functional, and Durable DNA Origami with Design-Specific Scaffolds." *ACS Nano* 13.5, pp. 5015–5027.

Felekyan, S., Kühnemuth, R., Kudryavtsev, V., Sandhagen, C., Becker, W., and Seidel, C. A. M. (2005). "Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection". *Review of Scientific Instruments* 76.8, p. 083104.

Felekyan, S., Kalinin, S., Sanabria, H., Valeri, A., and Seidel, C. A. M. (2012). "Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules." *Chemphyschem : a European journal of chemical physics and physical chemistry* 13.4, pp. 1036–1053.

Felekyan, S., Sanabria, H., Kalinin, S., Kühnemuth, R., and Seidel, C. A. M. (2013). "Analyzing Förster resonance energy transfer with fluctuation algorithms." *Methods in enzymology* 519, pp. 39–85.

Förster, T. (1948). "Zwischenmolekulare energiewanderung und fluoreszenz". *Annalen der Physik* 2, pp. 55–75.

Franck, J. and Dymond, E. G. (1926). "Elementary processes of photochemical reactions". *Transactions of the Faraday Society* 21, pp. 536–542.

Funke, J. J. and Dietz, H. (2016). "Placing molecules with Bohr radius resolution using DNA origami." *Nature nanotechnology* 11.1, pp. 47–52.

Ha, T., Enderle, T., Ogletree, D. F., Chemla, D. S., Selvin, P. R., and Weiss, S. (1996). "Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor". *Proceedings of the National Academy of Sciences* 93.13, pp. 6264–6268.

Hahn, J., Chou, L. Y. T., Sørensen, R. S., Guerra, R. M., and Shih, W. M. (2020). "Extrusion of RNA from a DNA-Origami-Based Nanofactory". *ACS Nano*.

Hartl, C., Frank, K., Amenitsch, H., Fischer, S., Liedl, T., and Nickel, B. (2018). "Position Accuracy of Gold Nanoparticles on DNA Origami Structures Studied with Small-Angle X-ray Scattering." *Nano Letters* 18.4, pp. 2609–2615.

He, Y., Lu, M., Cao, J., and Lu, H. P. (2012). "Manipulating protein conformations by single-molecule AFM-FRET nanoscopy." *ACS Nano* 6.2, pp. 1221–1229.

Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kühnemuth, R., Rezaei Adariani, S., Ambrose, B., Aznauryan, M., Barth, A., Birkedal, V., Bowen, M. E., Chen, H., Cordes, T., Eilert, T., Fijen, C., Gebhardt, C., Götz, M., Gouridis, G., Gratton, E., Ha, T., Hao, P., Hanke, C. A., Hartmann, A., Hendrix, J., Hildebrandt, L. L., Hirschfeld, V., Hohlbein, J., Hua, B., Hübner, C. G., Kallis, E., Kapanidis, A. N., Kim, J.-Y., Krainer, G., Lamb, D. C., Lee, N. K., Lemke, E. A., Levesque, B., Levitus, M., McCann, J. J., Naredi-Rainer, N., Nettels, D., Ngo, T., Qiu, R., Robb, N. C., Röcker, C., Sanabria, H., Schlierf, M., Schröder, T., Schuler, B., Seidel, H., Streit, L., Thurn, J., Tinnefeld, P., Tyagi, S., Vandenberk, N., Vera, A. M., Weninger, K. R., Wünsch, B., Yanez-Orozco, I. S., Michaelis, J., Seidel, C. A. M., Craggs, T. D., and Hugel, T. (2018). "Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study." *Nature methods* 15.9, pp. 669–676.

Hoeijmakers, J. H. J. (2009). "DNA damage, aging, and cancer." *The New England journal of medicine* 361.15, pp. 1475–1485.

Hübner, C. G., Renn, A., Renge, I., and Wild, U. P. (2001). "Direct observation of the triplet lifetime quenching of single dye molecules by molecular oxygen". *The Journal of Chemical Physics* 115.21, p. 9619.

Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., Portella, G., Battistini, F., Gelpi, J. L., González, C., Vendruscolo, M., Laughton, C. A., Harris, S. A., Case, D. A., and Orozco, M. (2016). "Parmbsc1: a refined force field for DNA simulations." *Nature methods* 13.1, pp. 55–58.

Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. M. (2012). "A toolkit and benchmark study for FRET-restrained high-precision structural modeling". *Nature methods* 9.12, pp. 1218–1225.

Kalinin, S., Valeri, A., Antonik, M., Felekyan, S., and Seidel, C. A. M. (2010). "Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states." *The journal of physical chemistry. B* 114.23, pp. 7983–7995.

Kapanidis, A. N., Laurence, T. A., Lee, N. K., Margeat, E., Kong, X., and Weiss, S. (2005). "Alternating-laser excitation of single molecules." *Accounts of chemical research* 38.7, pp. 523–533.

Kapusta, P. (2010). "Absolute Diffusion Coefficients: Compilation of Reference Data for FCS Calibration." *PicoQuant Application Note*.

Kasha, M. (1950). "Characterization of electronic transitions in complex molecules". *Discussions of the Faraday society*.

Khara, D. C., Schreck, J. S., Tomov, T. E., Berger, Y., Ouldridge, T. E., Doye, J. P. K., and Nir, E. (2018). "DNA bipedal motor walking dynamics: an experimental and theoretical study of the dependency on step size." *Nucleic acids research* 46.3, pp. 1553–1561.

Kopperger, E., List, J., Madhira, S., Rothfischer, F., Lamb, D. C., and Simmel, F. C. (2018). "A self-assembled nanoscale robotic arm controlled by electric fields." *Science* 359.6373, pp. 296–301.

Kudryavtsev, V., Sikor, M., Kalinin, S., Mokranjac, D., Seidel, C. A. M., and Lamb, D. C. (2012). "Combining MFD and PIE for accurate single-pair Förster resonance energy transfer measurements." *Chemphyschem : a European journal of chemical physics and physical chemistry* 13.4, pp. 1060–1078.

L Stryer, R. P. H. (1967). "Energy transfer: a spectroscopic ruler." *Proceedings of the National Academy of Sciences of the United States of America* 58.2, p. 719.

Lakowicz, J. R. (2006). *Principles of fluorescence spectroscopy. 3rd edn Springer*. New York.

Laurence, T. A., Kwon, Y., Yin, E., Hollars, C. W., Camarero, J. A., and Barsky, D. (2007). "Correlation spectroscopy of minor fluorescent species: signal purification and distribution analysis." *Biophysical journal* 92.6, pp. 2184–2198.

Lehman, I. R. (1974). "DNA ligase: structure, mechanism, and function." *Science* 186.4166, pp. 790–797.

Lerner, E., Cordes, T., Ingargiola, A., Alhadid, Y., Chung, S., Michalet, X., and Weiss, S. (2018). "Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer." *Science* 359.6373, eaan1133.

Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., and Webb, W. W. (2003). "Zero-mode waveguides for single-molecule analysis at high concentrations." *Science* 299.5607, pp. 682–686.

Lieber, M. R. (2010). "The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway". *Annual review of biochemistry* 79.1, pp. 181–211.

Limp-Foster, M. and Kelley, M. R. (2000). "DNA repair and gene therapy: implications for translational uses." *Environmental and molecular mutagenesis* 35.2, pp. 71–81.

Magde, D., Elson, E., and Webb, W. W. (1972). "Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy". *Physical Review Letters*.

Mari, P.-O., Florea, B. I., Persengiev, S. P., Verkaik, N. S., Brüggenwirth, H. T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J. A. A., Luider, T. M., Houtsmuller, A. B., and Gent, D. C. van (2006). "Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4." *Proceedings of the National Academy of Sciences* 103.49, pp. 18597–18602.

Maus, M., Cotlet, M., Hofkens, J., Gensch, T., De Schryver, F. C., Schaffer, J., and Seidel, C. A. (2001). "An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules." *Analytical chemistry* 73.9, pp. 2078–2086.

McKinney, S. A., Joo, C., and Ha, T. (2006). "Analysis of single-molecule FRET trajectories using hidden Markov modeling." *Biophysical journal* 91.5, pp. 1941–1951.

Milles, L. F., Schulten, K., Gaub, H. E., and Bernardi, R. C. (2018). "Molecular mechanism of extreme mechanostability in a pathogen adhesin." *Science* 359.6383, pp. 1527–1533.

Minsky, M. (1961). "Microscopy apparatus".

Müller, B. K., Zaychikov, E., Bräuchle, C., and Lamb, D. C. (2005). "Pulsed Interleaved Excitation". *Biophysical journal* 89.5, pp. 3508–3522.

Nettels, D., Gopich, I. V., Hoffmann, A., and Schuler, B. (2007). "Ultrafast dynamics of protein collapse from single-molecule photon statistics." *Proceedings of the National Academy of Sciences* 104.8, pp. 2655–2660.

Nickels, P. C., Wünsch, B., Holzmeister, P., Bae, W., Kneer, L. M., Grohmann, D., Tinnefeld, P., and Liedl, T. (2016). "Molecular force spectroscopy with a DNA origami-based nanoscopic force clamp." *Science* 354.6310, pp. 305–307.

Nicoli, F., Barth, A., Bae, W., Neukirchinger, F., Crevenna, A. H., Lamb, D. C., and Liedl, T. (2017). "Directional Photonic Wire Mediated by Homo-Förster Resonance Energy Transfer on a DNA Origami Platform." *ACS Nano* 11.11, pp. 11264–11272.

Nir, E., Michalet, X., Hamadani, K. M., Laurence, T. A., Neuhauser, D., Kovchegov, Y., and Weiss, S. (2006). "Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments." *The journal of physical chemistry. B* 110.44, pp. 22103–22124.

Otto, M. (2009). "Staphylococcus epidermidis–the 'accidental' pathogen." *Nature Reviews Microbiology* 7.8, pp. 555–567.

Pastwa, E., Somiari, R. I., Malinowski, M., Somiari, S. B., and Winters, T. A. (2009). "In vitro non-homologous DNA end joining assays–the 20th anniversary." *The international journal of biochemistry & cell biology* 41.6, pp. 1254–1260.

Piana, S., Donchev, A. G., Robustelli, P., and Shaw, D. E. (2015). "Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States". *The journal of physical chemistry. B* 119.16, pp. 5113–5123.

Ponnuraj, K., Bowden, M. G., Davis, S., Cell, S. G., and 2003 (2003). "A "dock, lock, and latch" structural model for a staphylococcal adhesin binding to fibrinogen". *Cell* 115, pp. 217–228.

Rabiner, L. R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE* 77.2, pp. 257–286.

Rayleigh, L. (1879). "XXXI. Investigations in optics, with special reference to the spectroscope". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*.

Reid, D. A., Keegan, S., Leo-Macias, A., Watanabe, G., Strande, N. T., Chang, H. H., Oksuz, B. A., Fenyo, D., Lieber, M. R., Ramsden, D. A., and Rothenberg, E. (2015). "Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair". *Proceedings of the National Academy of Sciences* 112.20, E2575–E2584.

Robustelli, P., Piana, S., and Shaw, D. E. (2018). "Developing a molecular dynamics force field for both folded and disordered protein states". *Proceedings of the National Academy of Sciences* 115.21, E4758–E4766.

Rothemund, P. W. K. (2006). "Folding DNA to create nanoscale shapes and patterns". *Nature* 440.7082, pp. 297–302.

Ruthardt, N., Lamb, D. C., and Bräuchle, C. (2011). "Single-particle tracking as a quantitative microscopy-based approach to unravel cell entry mechanisms of viruses and pharmaceutical nanoparticles." *Molecular therapy : the journal of the American Society of Gene Therapy* 19.7, pp. 1199–1211.

Saccà, B., Meyer, R., Erkelenz, M., Kiko, K., Arndt, A., Schroeder, H., Rabe, K. S., and Niemeyer, C. M. (2010). "Orthogonal protein decoration of DNA origami." *Angewandte Chemie International Edition* 49.49, pp. 9378–9383.

Salem, C. B., Ploetz, E., and Lamb, D. C. (2019). "Chapter 2 - Probing dynamics in single molecules". *Spectroscopy and Dynamics of Single Molecules*. Elsevier, pp. 71–115.

San Filippo, J., Sung, P., and Klein, H. (2008). "Mechanism of eukaryotic homologous recombination." *Annual review of biochemistry* 77.1, pp. 229–257.

Sánchez-Rico, C., Voith von Voithenberg, L., Warner, L., Lamb, D. C., and Sattler, M. (2017). "Effects of Fluorophore Attachment on Protein Conformation and Dynamics Studied by spFRET and NMR Spectroscopy". *Chemistry - A European Journal* 23.57, pp. 14267–14277.

Santiago-Frangos, A., Kavita, K., Schu, D. J., Gottesman, S., and Woodson, S. A. (2016). "C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA." *Proceedings of the National Academy of Sciences of the United States of America* 113.41, E6089–E6096.

Schrimpf, W., Barth, A., Hendrix, J., and Lamb, D. C. (2018). "PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data." *Biophysical journal* 114.7, pp. 1518–1528.

Schwille, P., Meyer-Almes, F. J., and Rigler, R. (1997). "Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution." *Biophysical journal* 72.4, pp. 1878–1886.

Selnihhin, D., Sparvath, S. M., Preus, S., Birkedal, V., and Andersen, E. S. (2018). "Multifluorophore DNA Origami Beacon as a Biosensing Platform." *ACS Nano* 12.6, pp. 5699–5708.

Shi, K., Bohl, T. E., Park, J., Zasada, A., Malik, S., Banerjee, S., Tran, V., Li, N., Yin, Z., Kurniawan, F., Orellana, K., and Aihara, H. (2018). "T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction." *Nucleic acids research* 46.19, pp. 10474–10488.

Shrestha, D., Jenei, A., Nagy, P., Vereb, G., and Szöllősi, J. (2015). "Understanding FRET as a Research Tool for Cellular Studies". *International Journal of Molecular Sciences* 16.4, pp. 6718–6756.

Sies, H. and Menck, C. F. (1992). "Singlet oxygen induced DNA damage." *Mutation research* 275.3-6, pp. 367–375.

Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Müller, S., and Seidel, C. A. M. (2011). "Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity". *Journal of the American Chemical Society* 133.8, pp. 2463–2480.

Stahl, E., Martin, T. G., Praetorius, F., and Dietz, H. (2014). "Facile and scalable preparation of pure and dense DNA origami solutions." *Angewandte Chemie International Edition* 53.47, pp. 12735–12740.

Stein, I. H., Capone, S., Smit, J. H., Baumann, F., Cordes, T., and Tinnefeld, P. (2012). "Linking Single-Molecule Blinking to Chromophore Structure and Redox Potentials". *Chemphyschem : a European journal of chemical physics and physical chemistry* 13.4, pp. 931–937.

Stein, I. H., Schüller, V., Böhm, P., Tinnefeld, P., and Liedl, T. (2011). "Single-molecule FRET ruler based on rigid DNA origami blocks." *Chemphyschem : a European journal of chemical physics and physical chemistry* 12.3, pp. 689–695.

Stokes, G. G. (1852). "On the change of refrangibility of light". *Phil Trans R Soc London* 142, pp. 463–562.

Taraporewala, Z. F., Jiang, X., Vasquez-Del Carpio, R., Jayaram, H., Prasad, B. V. V., and Patton, J. T. (2006). "Structure-function analysis of rotavirus NSP2 octamer by using a novel complementation system." *Journal of virology* 80.16, pp. 7984–7994.

Taraporewala, Z. F. and Patton, J. T. (2001). "Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2." *Journal of virology* 75.10, pp. 4519–4527.

Tomov, T. E., Tsukanov, R., Masoud, R., Liber, M., Plavner, N., and Nir, E. (2012). "Disentangling Subpopulations in Single-Molecule FRET and ALEX Experiments with Photon Distribution Analysis". *Biophysical journal* 102.5, pp. 1163–1173.

Trads, J. B., Tørring, T., and Gothelf, K. V. (2017). "Site-Selective Conjugation of Native Proteins with DNA." *Accounts of chemical research* 50.6, pp. 1367–1374.

Verlet, L. (1967). "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". *Physical Review* 159.1, pp. 98–103.

Viterbi, A. (1967). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm". *IEEE Transactions on Information Theory* 13.2, pp. 260–269.

Vogelsang, J., Kasper, R., Steinhauer, C., Person, B., Heilemann, M., Sauer, M., and Tinnefeld, P. (2008). "A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes." *Angewandte Chemie International Edition* 47.29, pp. 5465–5469.

Wagenbauer, K. F., Engelhardt, F. A. S., Stahl, E., Hechtl, V. K., Stömmer, P., Seebacher, F., Meregalli, L., Ketterer, P., Gerling, T., and Dietz, H. (2017). "How We Make DNA Origami." *ChemBioChem* 18.19, pp. 1873–1885.

Widengren, J., Kudryavtsev, V., Antonik, M., Berger, S., Gerken, M., and Seidel, C. A. M. (2006). "Single-molecule detection and identification of multiple species by multiparameter fluorescence detection." *Analytical chemistry* 78.6, pp. 2039–2050.

Yano, K.-i. and Chen, D. J. (2008). "Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway." *Cell cycle (Georgetown, Tex.)* 7.10, pp. 1321–1325.

Zarrabi, N., Schluesche, P., Meisterernst, M., Börsch, M., and Lamb, D. C. (2018). "Analyzing the Dynamics of Single TBP-DNA-NC2 Complexes Using Hidden Markov Models." *Biophysical journal* 115.12, pp. 2310–2326.

List of Figures

List of Abbreviations

2CDE	two-channel kernel density estimator
ACF	autocorrelation function
APD	avalanche photodiode
ALEX	alternating laser excitation
AOTF	acousto-optical tunable filter
AV	accessible volume
C-terminus	carboxy terminus (of a protein)
CCF	cross-correlation function
cryoEM	electron microscopy at cryogenic temperatures
CTR	C-terminal region
Cy3b	cyanine dye 3b
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DLL	dock, lock and latch
DSB	(DNA) double-strand break
E. coli	*Escherichia Coli*
EMCCD	electron multiplying charge coupled device
FC(C)S	fluorescence (cross-) correlation spectroscopy
Fgβ	fibrinogen β
FPGA	field-programmable gated array
FPS	FRET positioning and screening
FRET	Förster resonance energy transfer
GODCAT	glucose oxidase and catalase
HMM	hidden Markov model
HPLC	high performance liquid chromatography
HR	homologous recombination
IRF	instrument response function
NA	numerical aperture
NHEJ	non-homologous end-joining
NHS	*N*-hydroysuccinimide
nt	nucleotide
MATLAB	MATrix:LABoratory
MD	molecular dynamics
MFD	multiparameter fluorescence detection
N-terminus	amino terminus (of a protein)
NSP2-(ΔC)	(C-terminally truncated) non-structural protein 2
PCA	protocatechutic acid
PCD	protocatechute-3,4-dioxygenase
PDA	photon distribution analysis
PEG	polyethylene glycol
PDB	protein data bank
PAM	PIE analysis with Matlab
PIE	pulsed-interleaved excitation
PBS	phosphate buffered saline
PSF	point spread function
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RNase	ribonuclease
ROXS	reducing and oxidizing system

S. epidermis	*Staphylococcus epidermis*
SdrG	SD-repeat protein G
sm	single-molecule
sp	single-pair
TCEP	tris(2-carboxyethyl)phosphine
TCSPC	time-correlated single photon counting
TDP	transition density plot
TIRF(M)	total internal reflection fluorescence (microscopy)
Tracy	trace intensity analysis
Trolox	6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
UniProt	universal protein database
UV	ultraviolet
ZMW	zero-mode waveguide

Acknowledgements

Has it really been 4.5 years since I started my PhD back in May 2016? I guess time truly flies when you enjoy what you are doing and have great people supporting you.
Don, thank you so much for letting me be part of your research group and for all the trust you put in me and my work throughout the years. Thanks for giving me the freedom to grow as a scientist while always supporting me. I also appreciate your attempt to evoke my interest in American Football (but to be honest I still do not quite get the rules).
Christoph, thank you very much for your continuous interest in my research, your scientific input and for agreeing to be the second examiner of this thesis.
Karl and **Ralf**, I do not only want to thank you for being part of my PhD committee, but also for your continuous support over the years as members of my thesis advisory committee. I hope you enjoyed our annual meetings as much as I did. Thank you for all your input and ideas!
Philip, thank you for all the lively discussions during the Monday morning seminars and for being part of my PhD committee.
I also would like to thank **Prof. Achim Hartschuh** for joining my PhD committee.
Working in the lab and traveling to stimulating scientific conferences involved a lot of administrative work - My thanks go to **Silke Steger** and **Moritz** for guiding me through it.
I appreciate the networking opportunities and financial support provided by IMPRS, CeNS, NIM, CIPSM, SFB1032 and SFB1035. I especially want to thank **Hans-Joerg**, **Ingrid** and **Maxi** from the IMPRS coordination office and **Susanne** from the CeNS office for their great support.
Science is more fun when working as a team, so let's hear it for all the wonderful scientists I had the pleasure to collaborate with over the last years. For the DNA origami project I would like to thank **Alvaro** for kick-starting the project and for introducing me to the right people. **Mauricio**, thanks for your help with DNA origami design and handling, and thank you, **Tim**, for providing the required materials. **Melisande**, thanks for your support in the lab during your bachelor thesis. It was a pleasure being your supervisor. **Anders**, I want to thank you very much for contributing molecular dynamics simulations and for the many discussions we had about my data (literally day and night).
For the NSP2 project, I would like to thank all collaborators for their work which resulted in a great manuscipt. I have to especially thank **Alex**. You are not only a marvellous (please note the British spelling!) scientist who introduced me to the (sometimes weird) world of RNA, but you have also become a good friend. Our Friday evening brainstorming session with Valetin playing the accordion in the background is one of many treasured memories!
For the SdrG project I would like to thank **Lukas** as well as his PhD supervisor **Prof. Herrmann Gaub** for providing such an interesting biological system. It is quite funny that we initiated our collaboration at a conference in the mountains of Vermont although we were both working in Munich at that time. **Simon**, thank you very much for contributing a lot to the data analysis and for having become an expert on molecular dynamics simulations. I would also like to thank **Fabian** for his assistance with measurements during his internship and am happy that he decided to stay in our group as a master student.
I would like to thank all other former and current members of AK Lamb for the fun times we had inside and outside the lab: **Waldi, Wehne, Lena, Daniela, Sushi, Bässem, Philipp, Evelyn, Nader, Chen, Ivo, Maria, Ganesh, Vanessa, Adrian, Frank** and **Damir**. Thanks a lot **Sushi, Bässem, Evelyn** and **Vanessa** for everything TIRF-related, **Frank** for your help with the Labview control software and **Lena** and **Ganesh** for your support in the biolab. I also have to mention **Simon** again and thank you for being my lunch buddy and for always helping me finish the ridiculously huge amounts of food I brought to work.
Finally, I like to thank **my parents** and **my sister Lara** for supporting me since day 1 (okay, Lara since day 585, but that is just because she is younger). Danke, dass ihr immer für mich da seid! Last but not least, I would like to thank **Anders** again. You are my everything and keep inspiring me every day. I can't wait to see what the future holds for us ♡.

Curriculum Vitae

Education

05/2016 - 09/2020	PhD in Physical Chemistry Ludwig-Maximilians-Universität München, Germany Supervisor: Prof. Don C. Lamb
10/2013 - 03/2016	Studies in Biochemistry, Master of Science Ludwig-Maximilians-Universität München, Germany Final grade: 1.22
09/2015 - 03/2016	Master's thesis Subject: "Development of DNA Origami Structures for Application in Single-Molecule Fluorescence Studies" Supervisor: Prof. Don C. Lamb
10/2010 - 09/2013	Studies in Chemistry and Biochemistry, Bachelor of Science Ludwig-Maximilians-Universität München, Germany Final grade: 1.62
05/2013 - 09/2013	Bachelor's thesis Subject: "Direct Visualization of Monomer-by-Monomer Actin Filament Polymerization Using Zero-Mode Waveguides" Supervisor: Prof. Don C. Lamb
09/2001 - 07/2010	Secondary school, Moll-Gymnasium Mannheim, Germany

Professional Experience

05/2016 - 09/2020	PhD candidate in Prof. Don C. Lamb's lab Ludwig-Maximilians-Universität München, Germany
10/2014 - 12/2014	Research internship at Prof. Emma Sparr's lab Lund University, Sweden Subject: "Influence of gangliosides on the adsorption of monomeric α-synuclein to supported lipid bilayers – a QCM-D study"
04/2014 - 06/2014	Research internship at Prof. Christian Haass's lab Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Germany Subject: "What causes the TDP-43 knockout phenotype? – Identification of potential targets by immunohistochemistry"

Honors and Prizes

10/2019	CeNS travel award (oral presentation at the 64th BPS meeting in San Diego)
07/2017 and 12/2018	Poster prize at the 21st and 24th IMPRS seminar
12/2016	Prize for the Master's thesis of the Römer-Stiftung

Appendix A

Appended Papers

A.1 Paper 1: A DNA Origami Platform for Single-Pair Förster Resonance Energy Transfer Investigation of DNA-DNA Interactions and Ligation

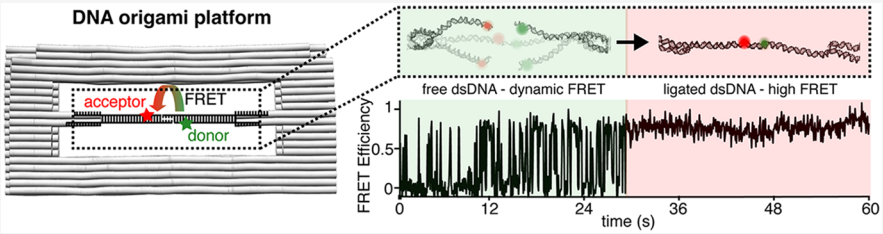
JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
Cite This: J. Am. Chem. Soc. 2020, 142, 815−825
pubs.acs.org/JACS

A DNA Origami Platform for Single-Pair Förster Resonance Energy Transfer Investigation of DNA−DNA Interactions and Ligation

Kira Bartnik,[†] Anders Barth,[†,§] Mauricio Pilo-Pais,[‡,⊥] Alvaro H. Crevenna,[†,∥] Tim Liedl,[‡] and Don C. Lamb*[,†]

[†]Department of Chemistry, Center for Nanoscience (CeNS), Nanosystems Initiative Munich (NIM) and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[‡]Department of Physics and Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität, 80539 Munich, Germany

Supporting Information



ABSTRACT: DNA double-strand breaks (DSBs) pose an everyday threat to the conservation of genetic information and therefore life itself. Several pathways have evolved to repair these cytotoxic lesions by rejoining broken ends, among them the nonhomologous end-joining mechanism that utilizes a DNA ligase. Here, we use a custom-designed DNA origami nanostructure as a model system to specifically mimic a DNA DSB, enabling us to study the end-joining of two fluorescently labeled DNA with the T4 DNA ligase on the single-molecule level. The ligation reaction is monitored by Förster resonance energy transfer (FRET) experiments both in solution and on surface-anchored origamis. Due to the modularity of DNA nanotechnology, DNA double strands with different complementary overhang lengths can be studied using the same DNA origami design. We show that the T4 DNA ligase repairs sticky ends more efficiently than blunt ends and that the ligation efficiency is influenced by both DNA sequence and the incubation conditions. Before ligation, dynamic fluctuations of the FRET signal are observed due to transient binding of the sticky overhangs. After ligation, the FRET signal becomes static. Thus, we can directly monitor the ligation reaction through the transition from dynamic to static FRET signals. Finally, we revert the ligation process using a restriction enzyme digestion and religate the resulting blunt ends. The here-presented DNA origami platform is thus suited to study complex multistep reactions occurring over several cycles of enzymatic treatment.

INTRODUCTION

Single-molecule Förster resonance energy transfer (FRET)[1] studies of biomolecules provide important insights into the dynamics of molecular interactions.[2] The efficiency of the energy transfer between a single donor and acceptor fluorophore pair is highly sensitive to their proximity,[3] enabling the measurement of inter- and intramolecular distances with sub-nanometer precision.[3,4] However, when using single-molecule FRET to measure interactions, the molecular concentrations necessary for weaker interactions are often too high to be investigated by single-molecule experiments. An elegant method to control the local concentration of the interaction partners is programmable DNA nanotechnology using self-assembled DNA origami.[5] These nanostructures consist of a long single-stranded scaffold and a set of short staple strands and allow the precise positioning of molecules such as fluorescent dyes,[6−8] proteins,[9] or DNA-based units[10] on the nanometer and even sub-nanometer scale.[11,12] DNA origamis are particularly suited for single-molecule studies, as a high local concentration of the molecules of interest can be achieved while maintaining a low enough total concentration for single-molecule detection. Thus, low-affinity interactions can be investigated and subpopulations or rare events identified, which would be otherwise hidden in ensemble measurements.[13] The combination of DNA nanotechnology with FRET offers a promising platform for the creation of advanced biosensors[14] and has provided key insights into the working principles of DNA nanomachines[10,15] or force-dependent DNA−protein interactions.[16]

Here, we use a DNA origami platform to specifically mimic a DNA double-strand break (DSB) and study the end-joining reaction on the single-molecule level using FRET. DNA DSBs are considered the most cytotoxic form of DNA damage that can lead to fragmentation or rearrangement of chromosomes when not properly repaired.[17] Since nonhomologous end-joining (NHEJ)[18] is one of the main repair pathways to restore genomic integrity in higher organisms, extensive studies have

Received: August 22, 2019
Published: December 4, 2019

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

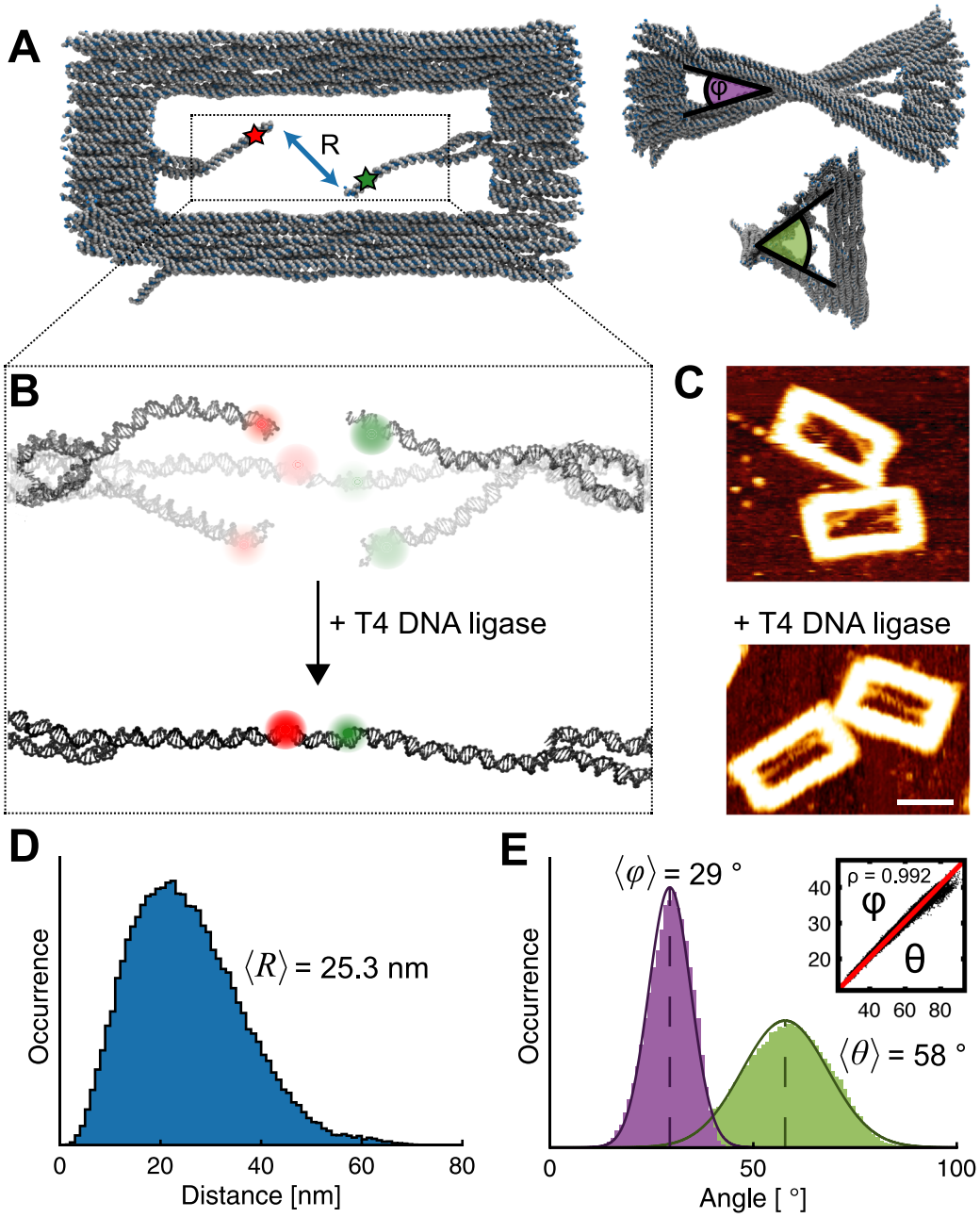


Figure 1. DNA origami platform for single-molecule detection of DNA ligation. (A) Schematic representations of the two-layer DNA origami frame with bound dsDNAs as repair substrates. The two double strands are labeled with Atto647N (red star) as FRET acceptor and Cy3b (green star) as FRET donor. The end-to-end distance R of the double strands is indicated by a blue arrow, and the twisting angles φ and θ are shown in purple and green, respectively. (B) DNA origami-bound double strands before (upper panel) and after ligation (lower panel). Before ligation, the double strands can move flexibly and the distance between the FRET dyes changes dynamically. After ligation, the interdye distance is fixed. (C) Exemplary AFM images for DNA origami platforms before (upper image) and after incubation with the T4 DNA ligase (lower image). For these experiments, double strands with a complementary overhang of 4 nucleotides were used. Scale bar: 50 nm. (D) End-to-end distance distribution of the origami frame with 4 nt overhangs obtained from coarse-grained MD simulations with an average distance of R = 25.3 ± 11 nm. (E) Distribution of the origami twisting angles obtained from coarse-grained MD simulations. φ and θ define the twisting between the longer sides and the shorter sides of the frame, respectively. The distributions are fitted to normal distributions, resulting in average angles of $\langle\varphi\rangle$ = 29 ± 5° and $\langle\theta\rangle$ = 58 ± 11. The angles φ and θ are highly correlated (see inset, Pearson correlation coefficient ρ = 0.992), indicating that the twisting motion globally affects the whole origami structure. The time evolution of the end-to-end distance and the twisting angles is shown in Figure S2 and illustrated in Movies S1 and S2.

been performed to understand the processes that occur during repair.[19−23] In a recent *in vitro* study, fluorescently labeled double-stranded DNAs (dsDNAs) were repaired by recombinant NHEJ proteins, and the mechanism was investigated by single-pair FRET (spFRET).[24] This experimental approach allowed new insights into the kinetics of the end-joining process and proposed a two-step mechanism of DNA end-pairing followed by an alignment within the synaptic complex. However, it relied on the random pairing of a freely diffusing dsDNA with a surface-immobilized dsDNA, thus not

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

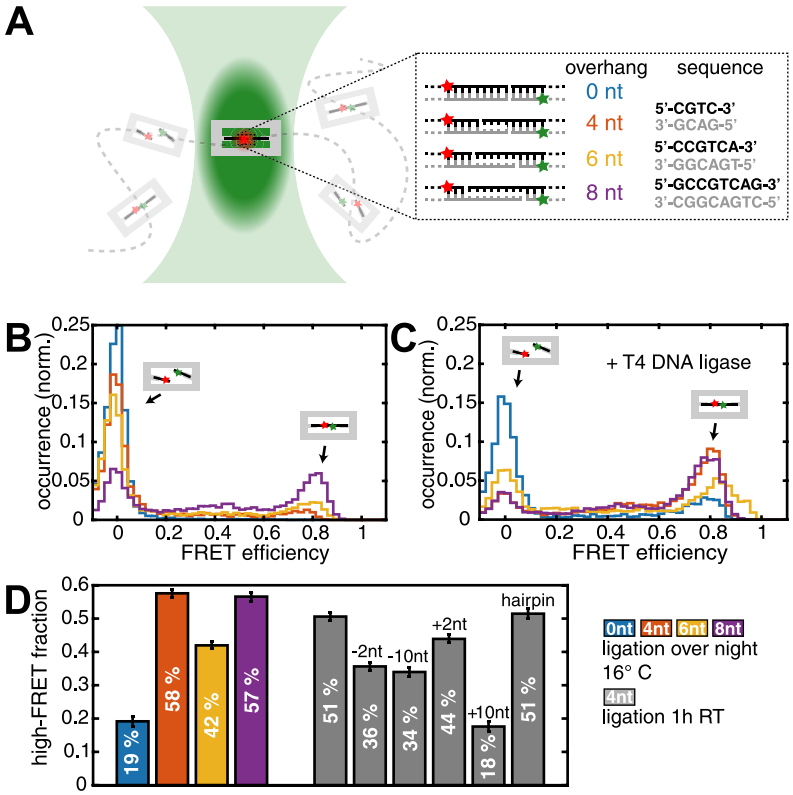


Figure 2. Detection of DNA ligation in solution. (A) DNA origami platforms with bound fluorescently labeled DNA double strands diffuse through the confocal volume. The complementary overhang length between the DNA double strands varies from 0 nt to 4 nt, 6 nt, and 8 nt (zoom-in; the red and green stars represent the labeling positions of the Atto647N FRET acceptor and Cy3b FRET donor dyes, respectively). The sequence is given for every overhang length. (B) Histograms of the molecule-wise spFRET efficiency before the addition of the T4 DNA ligase. (C) Histograms of the spFRET efficiency after incubation with the T4 DNA ligase. (D) Comparison of the high-FRET efficiency fractions (FRET efficiency E larger than 0.6) after ligation for different overhang lengths, incubation conditions, and overall double-strand length indicates the most successful ligation reaction for 4 nt overhangs on regular length DNA double strands after incubation overnight at 16 °C. The exact numbers of high-FRET and the total amount of molecules are listed in Table S2. Error bars represent 95% confidence intervals.

accounting for the high local concentration of broken DNA ends in the cell.

To this end, inspired by similar designs,[25−27] we designed a DNA origami frame using a 7560-nucleotide (nt) scaffold that positions two DNA double strands in a linear geometry to monitor the repair process by spFRET (Figure 1). To catalyze the end-joining of the fluorescently labeled DNA duplexes within the DNA origami gap, we chose the DNA ligase from bacteriophage T4 (T4 DNA ligase) as a model enzyme for repair, which is widely used in molecular biology.[28] Similar to the DNA ligase IV involved in NHEJ, this enzyme requires juxtaposed 5′-phosphate and 3′-hydroxyl termini for successful ligation.[29,30] The ligation process was studied for freely diffusing DNA origami structures as well as for surface-immobilized platforms. The modularity of DNA origami allowed us to compare the ligation efficiencies of dsDNAs with different lengths of complementary overhangs and sequences. The DNA origami design not only enables us to follow the ligation process in real time but makes it also possible to ligate and cut the same DNA double-strand pairs repeatedly, which would not be possible for a mixture of freely diffusing DNA strands. Our results demonstrate the applicability of our DNA origami platform to a wide range of multimolecular interaction studies.

RESULTS AND DISCUSSION

A DNA Origami Platform for Monitoring DNA Ligation. A schematic of the DNA origami platform is shown in Figure 1A. The two-layer DNA origami frame harbors two protrusions inside its gap, which serve as attachment sites for the molecules of interest. Here, two approximately 80-nt-long DNA double strands are bound within the origami as DNA repair substrates (the length of the different sequences used are given in Figure 2A and Table S1). To monitor the repair process, one double strand is fluorescently labeled with Cy3b as a FRET donor and the other one with Atto647N as a FRET acceptor. Before being joined to one continuous strand, the two individual DNA double strands are expected to move flexibly inside the origami gap with the possibility of transient binding of the complementary overhangs (Figure 1B, top). While the free DNA is expected to exhibit a dynamic FRET signal, a static FRET signal should be observed after DNA repair due to the fixed distance between the dyes (Figure 1B, bottom). dsDNA has been extensively studied using spFRET and has recently been benchmarked in a seminal study by 20 laboratories.[4]

The correct folding of the DNA origami frame was confirmed using atomic force microscopy (AFM, Figure S1). Imaging of the attached DNA double strands before and after incubation with the T4 DNA ligase indicated a successful end-joining reaction on the origami platform. Before ligation, the

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

two DNA double strands were randomly positioned within the origami gap, while incubation with the T4 DNA ligase connected them to one strand (Figure 1C). To characterize the intrinsic dynamics of the origami platform, coarse-grained molecular dynamics (MD) simulations based on the oxDNA model[31] were performed. In the simulations, we observed unhindered movement of the unligated DNA double strands and a twisting motion of the origami frame with an average twisting angle of 29° along the long axis and 58° along the short axis (Figure 1D and E, Figures S2 and S3, and Supplementary Movies S1 and S2).

Ligation in Solution. To test if we could monitor DNA ligation using the DNA origami platform by spFRET, DNA double strands with different overhang lengths were bound to the DNA origami. Complementary overhangs of 0 nt (blunt ends) and 4, 6, and 8 nt (sticky ends) were investigated (Figure 2A, zoom-in). The positions of the dyes (and thus the interdye distance) as well as the DNA sequence were kept constant between these constructs (the sequence and labeling positions are given in Table S1). We monitored freely diffusing DNA origami platforms in solution using a confocal microscope with multiparameter fluorescence detection (MFD) and pulsed interleaved excitation (PIE).[32,33] Single DNA origami platforms diffused through the femtoliter-sized observation volume on the time scale of 30 ms (Figure S4), resulting in bursts of photons, enabling us to measure the FRET efficiency of individual molecules.

Before the addition of a repair enzyme, the origami-bound DNA double strands showed a predominant low-FRET population at a FRET efficiency E around 0 for the various overhang lengths (Figure 2B). In this case, the termini of the two fluorescently labeled DNA double strands are too far apart for FRET between the Cy3b donor and the Atto647N acceptor. For the blunt-end construct, the low-FRET population was the only population detected, while the sticky-end constructs showed an additional high-FRET population at a FRET efficiency E around 0.8. When the sticky ends bind, the donor−acceptor FRET pair is separated by 12 nt. From our experimental data, we determined an interdye distance of 52.9 Å by photon distribution analysis (PDA).[34] Based on calculations of the accessible volume of the dyes[35,36] (Figure S5A), the expected FRET-averaged distance on the ligated DNA is 47.9 Å (corresponding to a FRET efficiency of 0.88), which is in good agreement with the experimental value. Hence, the high-FRET population indicates transient binding of the sticky ends to each other. The longer the complementary overhang length, the higher the population of molecules that exhibit the high-FRET state, while the low-FRET population decreases with increasing overhang length (Figure 2B). To characterize the free state before enzyme addition, we simulated the movement of DNA origami-bound double strands with a 4-nt overhang and extracted a mean distance between the sticky ends of 25.3 ± 11.0 nm (Figure 1D), which is indeed too far apart for efficient energy transfer between the fluorophores. Furthermore, our coarse-grained MD simulations showed distance fluctuations between 5 and 60 nm and therefore fluctuations in FRET efficiency, similar to our experimental findings. However, no hybridization of the sticky ends was observed in the simulations.

To connect the two DNA double strands, we used the T4 DNA ligase as a model enzyme, which catalyzes the formation of a phosphodiester bond between proximate 5′-phosphate and 3′-hydroxyl ends of nucleic acids,[29] and a buffer containing 1 mM ATP. This enzyme has been recently used to enhance the stability of DNA origami nanostructures that contained phosphorylated staple strands with no detrimental effect on the overall structural integrity.[37] Thus, the T4 DNA ligase is not expected to compromise the structure of our origami frame. Further, the ligation reaction specifically targets the fluorescently labeled DNA double strands as they carry a 5′-phosphate group, which the DNA origami staple strands lack. Once ligated, a high-FRET efficiency state is expected. Indeed, incubation of the DNA origami-bound double strands with the T4 DNA ligase resulted in a decrease of the low-FRET population at $E = 0$ (corresponding to unligated dsDNA) and an increase in the high-FRET population at $E = 0.8$ (corresponding to ligated dsDNA) for all overhang lengths (Figure 2C).

We further observed an intermediate-FRET state, which was also present for 8-nt-sticky-end constructs before ligation. This state is most likely caused by misfolded structures or small aggregates, as it was more pronounced for DNA origami platforms that were not purified by agarose gel electrophoresis. The number of these aggregates was slightly increased in the presence of the T4 DNA ligase. However, we also observed the intermediate-FRET state for isolated control DNA double strands without the origami platform after ligation (Figure S6), indicating that photophysical artifacts may contribute to the occurrence of this state. As the sequence between the FRET dyes is identical for the different constructs, the repair process is expected to result in the same continuous DNA double strand and therefore the same FRET efficiency, regardless of the overhang length. While the same FRET efficiency was observed for the high-FRET population, the ligation efficiency is clearly different. The blunt-end construct showed only 19% single-molecule events with $E > 0.6$, while sticky-end double-stranded DNAs led to at least 42% high-FRET efficiency events (Figure 2D). Thus, the sticky ends were ligated more efficiently by the enzyme than the blunt ends under our experimental conditions. While incubation with the T4 DNA ligase resulted in almost identical FRET efficiency histograms for 4- and 8-nt overhangs (Figure 2C) and thus similar fractions of high-FRET efficiency events of 58% and 57%, respectively (orange and purple bars in Figure 2D), only 42% of the population were observed in the high-FRET efficiency state for the 6-nt construct (yellow bar in Figure 2D). Possible explanations for the lower ligation efficiency are the formation of competing secondary structures within this specific single-stranded overhang,[38] binding of the overhang to complementary and unoccupied parts of the DNA origami scaffold strand, which was occasionally observed by AFM (Figure 1C, upper panel), or missing 5′-phosphate groups.

To address whether the DNA origami platform has an influence on ligation efficiency, we performed control experiments on freely diffusing 4-nt overhang DNA double strands. For these experiments, the single-stranded overhangs used for attachment to the origami platform were deleted, leaving blunt ends on the opposite side. Under the same experimental conditions, ligation was less efficient for the control dsDNAs without the origami, as a lower fraction of high-FRET efficiency events was observed compared to 4-nt overhang DNA strands bound to our platform. However, significantly increasing the concentration of the isolated dsDNA as well as the concentration of the T4 DNA ligase resulted in similar high-FRET fractions with and without the DNA origami frame

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

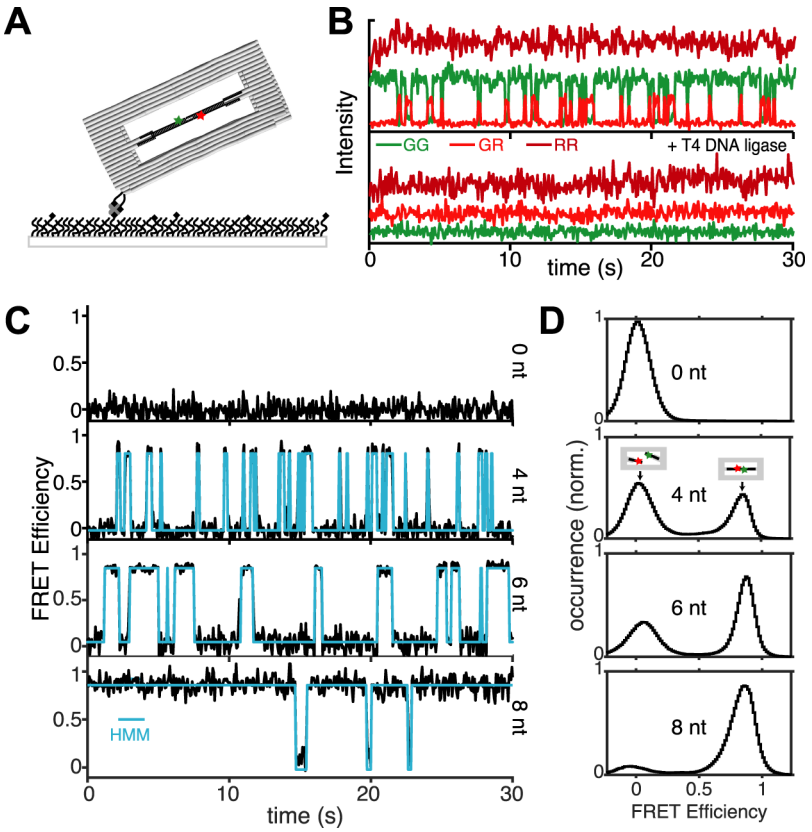


Figure 3. Detection of DNA dynamics and ligation on surface-immobilized DNA origamis. (A) DNA origamis are anchored via two biotinylated staples to a biotinylated PEG surface using streptavidin and illuminated by TIRF excitation. (B) Exemplary fluorescence intensity traces of origami-bound DNA double strands with 4-nt overhangs before (upper panel) and after incubation with the T4 DNA ligase (lower panel). The donor signal after donor excitation is shown in green, the acceptor signal after donor excitation (i.e., the FRET signal) is shown in red, and the acceptor signal after acceptor excitation (i.e., the ALEX signal) is shown in dark red. (C) Exemplary FRET efficiency traces for the different overhang lengths before incubation with the T4 DNA ligase. The FRET efficiency traces (in black) are analyzed using a two-state hidden Markov model (HMM, in cyan). (D) Framewise histograms over all recorded FRET traces show an increase in the high-FRET population with increasing overhang length. The numbers of analyzed molecules (and corresponding total transitions between low- and high-FRET efficiency values) are: 1151 molecules (0 transitions) for 0-nt, 814 molecules (190 985 transitions) for 4-nt, 556 molecules (79 046 transitions) for 6-nt, and 489 molecules (5699 transitions) for 8-nt complementary overhangs.

(Figure S6 and Table S3). Thus, the locally high DNA concentration on the origami has a positive effect on ligation efficiency.

For all tested overhang lengths, incubation with the T4 DNA ligase considerably reduced the low-FRET population and therefore the amount of unligated dsDNA, but ligation was never complete. To exclude geometrical constraints as a reason for the incomplete ligation, we varied the length of the DNA repair substrates. In the previous experiments, the dsDNA sequences were designed to exactly span the gap within the DNA origami frame once the two dsDNA ends were ligated together. We repeated experiments using a 4-nt overhang with DNA sequences that were either shorter or longer than the original design (sequences are given in Table S1). Before ligation, DNAs with an addition or a gap of 2 nt showed a similar behavior to the regular length DNAs with two subpopulations, a major low-FRET population and a minor high-FRET population, indicative of transient binding. To provide even more flexibility, we additionally incorporated two hairpin motifs into one DNA double strand, which led to a similar result. The addition or removal of 10 nucleotides, however, hindered the transient binding of the overhangs, as no high-FRET efficiency events were observed (Figure S7A). After incubation with the T4 DNA ligase, dsDNA with 2 additional nucleotides or incorporated hairpin sequences were almost as efficiently repaired as the regular length, as they showed similar high-FRET efficiency fractions of 44%, 51%, and 51%, respectively (Figure 2D). A gap of 2 or 10 nucleotides decreased the ligation efficiency slightly to 36% and 34%, respectively, while an addition of 10 nucleotides resulted in only 18% high-FRET efficiency events (Figures 2D and S7B). Overall, shortening or elongation of the DNA double strands could not increase the ligation efficiency above the already observed level. Thus, the length of the repair substrates does not seem to be the reason for an incomplete ligation reaction. Moreover, the successful ligation of the shortened constructs supports the MD finding that the DNA origami frame is indeed flexible.

Next, we looked at the influence of incubation time and temperature on ligation efficiency. Incubation overnight at 16 °C resulted in a more pronounced high-FRET population and therefore more ligated dsDNA compared to an incubation at room temperature for 1 h for 4-nt-sticky-end constructs (Figure 2D, 58% and 51%, respectively, and Figure S8). In another single-molecule study, the temperature-dependence of the T4 DNA ligase was already observed for 25 and 37 °C. Also here, ligation was never complete and the yield of ligation products could not exceed 88 ± 1% (for 18 h at 25 °C).[39] This

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

value is higher than our results. However, we reason that an imperfect ligation reaction is mainly linked to the incubation conditions as well as the enzymatic activity and not directly connected to the DNA origami design.

Ligation with Immobilized Structures. To follow the ligation of individual DSBs over the time scale of seconds to minutes, we measured surface-immobilized DNA origami platforms using total internal reflection fluorescence microscopy (TIRFM). For surface anchoring, two DNA staples were elongated with a T_5-spacer and biotin in one corner of the origami, which allowed specific binding of the origamis to a biotinylated surface via streptavidin (Figure 3A). Initially, we had attempted to anchor the platform on all four corners, but observed that this approach hindered the intrinsic dynamics of the platform. The fluorescently labeled DNA double strands could not dissociate, and it was not possible to observe the ligation process with this anchoring strategy. We applied millisecond alternating laser excitation (ALEX)[40] to confirm the presence of both the Cy3b donor and Atto647N acceptor dyes and to exclude photophysical effects such as acceptor dark states as a reason for dynamic FRET changes.

DNA origami-bound double strands were measured before enzyme treatment and after incubation with the T4 DNA ligase. A typical fluorescence intensity time trace for dsDNAs with 4-nt complementary overhangs is shown in Figure 3B (more examples are shown in Figure S9B). Before ligation, anticorrelated donor and acceptor signals were observed. A high donor fluorescence intensity (in green) indicates unbound DNA double strands and a high acceptor fluorescence intensity (in red) transient binding of the 4-nt overhangs. The direct probing of the acceptor (ALEX trace, dark red) verifies that photophysical artifacts of the acceptor are not contributing to the dynamics. After incubation with the T4 DNA ligase, these dynamic intensity fluctuations disappeared in favor of a continuous high acceptor intensity, which suggests ligated DNA double strands.

Exemplary spFRET efficiency traces for the different overhang lengths before ligation are shown in Figures 3C and S9. Similar to the spFRET experiments in solution (Figure 2B), blunt-end constructs showed no FRET at all, while sticky-end constructs displayed dynamic fluctuations between a low-FRET efficiency state at $E = 0$ (corresponding to unbound DNA double strands) and a high-FRET efficiency state around $E = 0.85$ (corresponding to transiently bound DNA double strands). To quantify the transition rates, we used a two-state hidden Markov model (HMM) analysis[41−43] to extract the mean dwell times of the high- and low-FRET states. For overhangs of 8 nt, only a few molecules displayed FRET transitions, while the majority showed a continuous high-FRET signal. Thus, only the rates for 4- and 6-nt overhangs could be reliably calculated. The transition rate from the high- to the low-FRET state decreased from $2.9 \pm 0.4\ s^{-1}$ for overhangs of 4 nt to $1.2 \pm 0.1\ s^{-1}$ for 6 nt (Figure S10A). This confirms the expected trend that longer complementary sequences increase the time scale of transient binding. The transition rates from the low- to the high-FRET state were almost identical, with $1.5 \pm 0.2\ s^{-1}$ for 4-nt and $1.2 \pm 0.1\ s^{-1}$ for 6-nt sticky ends (Figure S10B). Thus, the association rate seems to be mostly independent of the overhang length.

To compare the different overhang lengths, we generated framewise histograms over all recorded FRET traces (Figure 3D). While blunt-end constructs showed only a single low-FRET population, sticky-end constructs displayed both a low- and a high-FRET population. As mentioned before, 8-nt complementary overhangs mostly resulted in continuous high-FRET signals without transitions, and the respective histogram thus shows a dominant high-FRET peak. As a general observation, an increase in overhang length shifts the equilibrium toward the high-FRET population and therefore the transiently bound state, as expected from the extracted transition rates.

After incubation with the T4 DNA ligase, continuous high-FRET traces were observed for all overhang lengths, indicating repaired and covalently bound DNA double strands (Figure S11). For blunt-end DNAs, only a few of these high-FRET traces were found, while most of the molecules still showed no FRET. This is in agreement with the solution spFRET data, which showed only a minor high-FRET population after ligation (Figure 2C). For 4- and 8-nt overhangs, the dynamic FRET signals completely disappeared in favor of the static high-FRET signal. For 6-nt sticky ends, however, incubation with the T4 DNA ligase led to both static high-FRET and dynamic FRET signals. Thus, some molecules still showed transient binding of the fluorescently labeled DNA double strands, indicating that they were not covalently linked. This result reproduces the findings of the spFRET measurements in solution that the specific sequence used for the 6-nt sticky ends is less efficiently repaired by the T4 DNA ligase compared to the other sticky-end constructs.

To follow the ligation process in real time, we immobilized origami-bound DNA double strands with 4-nt overhangs on the surface and added a low concentration of the T4 DNA ligase (around 100 pM). While many molecules showed either dynamic fluctuations of the FRET efficiency or a continuous high-FRET signal, some molecules exhibited a transition from the dynamic behavior to a static, high-FRET state (Figure 4).

For example, in the beginning of the FRET efficiency trace shown in Figure 4A, fast fluctuations between the low- and high-FRET state are observed. After 30 s, the dynamics stop and a continuous high-FRET signal appears. To exclude a photophysical origin of this change, the acceptor signal after acceptor excitation is plotted as well (dark red curve in Figure 4A). As the direct acceptor signal is stable over time, changes in FRET efficiency have to be caused by an interaction of the DNA double strands with the T4 DNA ligase. Some additional examples for the ligation reaction are shown in Figure 4B. Thus, our DNA origami platform enables us to follow the ligation of two individual DNA double strands in real time.

Reversibility of the Ligation Process. To test the reversibility of the ligation process, we modified the DNA sequences of the 4-nt sticky-end construct by introducing the recognition site of the restriction enzyme ZraI, 5′-GACGTC-3′, positioned between the dyes. This enzyme was chosen as its recognition sequence is not part of the DNA origami sequence. Thus, restriction digestion specifically targets the fluorescently labeled DNA double strands. We first incubated the origami platforms with the T4 DNA ligase, then cut with the restriction enzyme ZraI, and finally religated using the T4 DNA ligase. By observing the increase or decrease of the high-FRET population, the ligation and cleavage of the DNA double strands could be monitored in solution (Figure 5).

Before addition of the T4 ligase, 22% of all molecules gave rise to a minor high-FRET population (FRET efficiency $E > 0.5$) indicative of transient binding due to the 4-nt complementary overhangs. This population is shifted to $E = 0.65$, which is a slightly lower FRET efficiency value compared

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

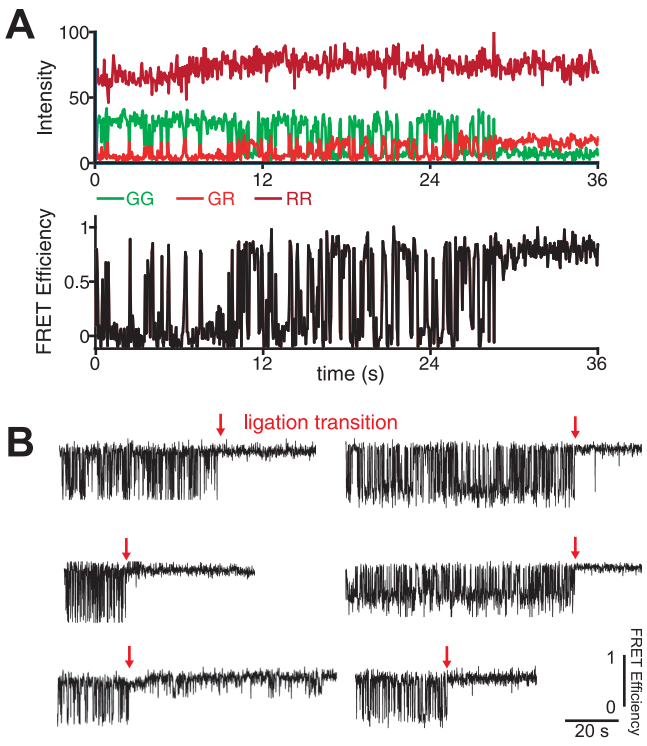


Figure 4. Observation of the ligation process by TIRF. (A) Exemplary FRET trace of origami-bound DNA double strands with 4-nt overhangs after addition of the T4 DNA ligase. The donor signal after donor excitation is shown in green, the acceptor signal after donor excitation (i.e., the FRET signal) is shown in red, and the acceptor signal after acceptor excitation (i.e., the ALEX signal) is shown in dark red. (B) Additional example traces show the transition from a dynamic to a continuous FRET signal.

to the DNA sequences without ZraI recognition sequence at E = 0.8 (Figure 2B), although the FRET pairs are separated by 12 nt in both cases. The difference in FRET efficiencies is caused by altered positions of the dyes on the DNA double helix (Figure S5), and the experimentally determined interdye distance of 59.7 Å matches the theoretical value of 57.2 Å (corresponding to a FRET efficiency of 0.72). Therefore, we chose a different threshold (E > 0.5) to calculate the high-FRET fraction compared to DNA sequences without ZraI recognition sequence (E > 0.6, Figure 2D).

After incubation with the T4 DNA ligase, the fraction of molecules in the high-FRET state increased to 39%, indicating successful ligation. Interestingly, ligation is not as efficient as for the 4-nt overhang dsDNAs without the restriction enzyme recognition site. Under the same experimental conditions, 58% of these molecules showed a high-FRET signal after ligation (Figure 2D). This finding suggests a sequence-specific preference of the T4 DNA ligase, which was described in detail for dsDNA with 3-nt overhangs.[39]

After a washing step to remove the T4 DNA ligase as well as EDTA from the buffer to ensure the functionality of the restriction enzyme, the DNA origami platforms were incubated with ZraI. This enzyme is expected to cut the ligated DNA double strands into blunt ends, therefore leading to a major low-FRET population. Indeed, the number of molecules displaying high-FRET decreased to 31%. However, a much lower value was expected, as blunt-end constructs should not show transient binding and therefore no high-FRET population at all. This indicates that not all ligated dsDNAs were cut, which is most likely due to nonoptimal reaction conditions. While an incubation temperature of 37 °C is recommended for ZraI digestion, we performed experiments at room temperature to avoid possibly damaging the DNA origami structures.

After an additional washing step to remove ZraI, we incubated the origami structures again with the T4 DNA ligase. The fraction of high-FRET molecules increased to 39%, which is identical to the result of the first ligation step. The DNA double strands, which have been cut by ZraI before, appear to be rejoined after the additional ligation step.

To verify that our DNA origami design makes it possible to ligate and cut the same pair of DNA double strands, we

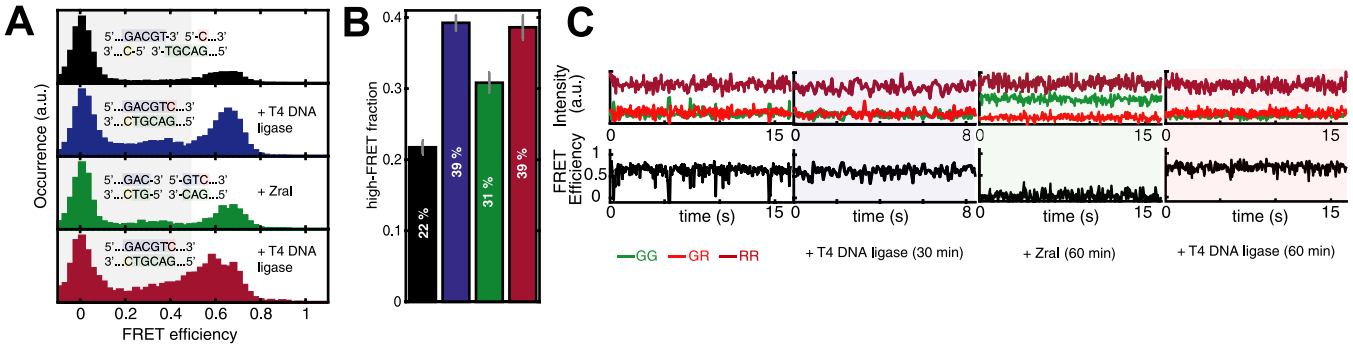


Figure 5. DNA ligation and subsequent cleavage in solution (A and B) and for surface-immobilized DNA origamis (C). (A) From top to bottom: Histograms of the spFRET efficiency before the addition of the T4 DNA ligase, after incubation with the T4 DNA ligase, after subsequent incubation with the restriction enzyme ZraI, and after another incubation step with the T4 DNA ligase. The expected sequences of the DNA double strands—starting with a complementary overhang of 4 nucleotides—are shown for each histogram. (B) Comparison of the high-FRET efficiency fractions (FRET efficiency E larger than 0.5) for the histograms in A. The exact numbers of low-FRET and the total amount of molecules are listed in Table S4. Error bars represent 95% confidence intervals. (C) Exemplary fluorescence intensity traces of origami-bound DNA double strands before the addition of the T4 DNA ligase, after incubation with the T4 DNA ligase, after subsequent incubation with the restriction enzyme ZraI, and after another incubation step with the T4 DNA ligase. The donor signal after donor excitation is shown in green, the acceptor signal after donor excitation (i.e., the FRET signal) is shown in red, and the acceptor signal after acceptor excitation (i.e., the ALEX signal) is shown in dark red. To record the trace, surface-bound DNA origami platforms were measured using TIRFM within a fixed field of view (Figure S12; the presented fluorescent intensity traces correspond to molecule 1).

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

immobilized the platforms and measured them using TIRFM before and after incubation with the T4 DNA ligase or ZraI. The field of view was fixed during the experiment in order to image the same set of molecules over the course of the enzyme reactions (Figure S12). In Figure 5C, we observe anticorrelated donor and acceptor signals before ligation due to transient binding of the sticky ends. After incubation with the T4 DNA ligase, the FRET signal becomes static with a high FRETefficiency indicating a successful ligation reaction. After incubation with the restriction enzyme ZraI, no FRET is observed, implying digestion of the ligated DNA into blunt ends. After a subsequent incubation with the T4 DNA ligase for 60 min, the molecule goes back to a static high-FRET signal and therefore to the ligated state. We observed several other origamis where the DNA double strands went from a dynamic to a static to a no-FRET state after the consecutive incubations with the T4 DNA ligase and ZraI. However, as ligation of blunt ends is less efficient than for sticky ends, we rarely observed FRET after the second ligation reaction (Figure S13). In addition, the temperature conditions and the limited incubation time on the microscope led to a further decrease in the expected yield. Thus, our DNA origami design allowed us to specifically ligate and cut DNA double strands bound to the origami, while the platform stayed intact over several cycles of enzyme treatment.

CONCLUSION

Here, we introduced a DNA origami platform for spFRET-based studies of DNA end-joining and ligation. As the origami design brings the two DNA double strands of interest close together, transient binding of the DNA double strands could be detected before the addition of any repair enzyme. The kinetics for several thousands of binding events were characterized and showed that the dissociation rates are affected by the length of the complementary overhang between the dsDNAs. As a model enzyme for the end-joining reaction, we used the DNA ligase from bacteriophage T4. Our spFRET experiments indicated a successful ligation reaction both in solution and for surface-anchored DNA origami platforms. While sticky ends were more efficiently ligated than blunt ends, we also detected differences in ligation efficiency for individual sticky-end lengths, sequences, and ligation reaction conditions. Using surface-tethered DNA origamis and DNA with 4-nt complementary overhangs, the ligation process could be observed directly. We could not only rejoin DNA ends but were also able to specifically cut double strands with a restriction enzyme. As a next step, the platform could be applied to study the interaction of complementary DNA sequences of different lengths or with incorporated mismatches, or to investigate the sequence specificity of the T4 DNA ligase. Further, it would be of high interest to unravel the complex dynamics occurring during NHEJ in higher organisms by using purified proteins on our platform.

In conclusion, our DNA origami provides an ideal platform for interaction studies of DNA-binding enzymes and fluorescently labeled DNA double strands on the single-molecule level. Our design can be easily adapted for various DNA/DNA, DNA/protein, and even protein/protein interaction studies. By cross-linking DNA sequences to proteins via various chemistries,[44] they can be attached to the DNA origami scaffold. As the stoichiometry of interaction partners can be specifically controlled, potential applications of our platform could be the investigation of early stages of actin nucleation or aggregation of α-synuclein monomers. Due to the induced proximity on our origami nanostructure, it will be possible to study weak interactions, which would otherwise be difficult to observe on the single-molecule level.

MATERIALS AND METHODS

DNA Origami Assembly and Purification. The DNA origami two-layer frame is based on the square lattice design with p7560 as the scaffold strand (for design details and sequences see Table S5). For the design, we were inspired by a similar DNA origami frame, which was used to study DNA base-excision repair by AFM.[25] Scaffold and staple strands were purchased from Eurofins Genomics, Ebersberg, Germany. The DNA repair substrates with 0-, 4-, and 8-nt overhangs were ordered from Purimex, Grebenstein, Germany. The DNA oligonucleotides with 6-nt overhangs as well as 4-nt overhangs with longer or shorter sequences (+2/+10/−2/−10 nt/hairpin) or restriction enzyme recognition sites (RE) were produced by IBA GmbH, Göttingen, Germany (an overview of the sequences can be found in Table S1).

To form dye-functionalized double-stranded DNAs, complementary DNA strands (see Figure S14 for a schematic) were annealed in 10 mM Tris, 1 mM EDTA, and 12 mM $MgCl_2$ at concentrations of 10 μM per oligo by heating to 99 °C for 5 min and then cooling to 9 °C at a rate of −1 °C/min.

The DNA origami was folded by mixing 10 nM scaffold with a 10-fold excess of staple strands in 10 mM Tris, 1 mM EDTA, and 12 mM $MgCl_2$. The mixture was first heated to 65 °C for 1 h and then cooled to 20 °C at a rate of −1 °C/h. Unbound staple strands were removed by two rounds of poly(ethylene glycol) precipitation.[45] The purified DNA origami structures were incubated with a 10-fold excess of Cy3b- and Atto647N-labeled DNA double strands at 20 °C under shaking at 300 rpm overnight.

To remove excess fluorescently labeled dsDNA and DNA origami aggregates, samples were separated on a 1% agarose gel containing 1× TAE buffer (40 mM Tris, 0.1% (v/v) acetic acid, 1 mM EDTA) and 11 mM $MgCl_2$ for 90 min at 80 V, while the gel box was cooled in an ice water bath. For imaging, the gel was incubated in a 1× solution of SYBR gold nucleic acid gel stain (ThermoFisher Scientific, Waltham, MA, USA) for 40 min. When samples were recovered from the gel, the staining step with the DNA intercalator was skipped and the Cy3b fluorescence was used instead to identify the correct DNA origami band. Gel extraction was performed with Freeze 'n Squeeze DNA gel extraction spin columns (Bio-Rad). Afterward, the samples were concentrated using Amicon Ultra centrifugal filters, 100 kDa MWCO (Merck, Darmstadt, Germany) using a buffer of 10 mM Tris, 1 mM EDTA, and 12 mM $MgCl_2$.

Single-Pair FRET Measurements in Solution. Single-pair FRET measurements of freely diffusing DNA origami platforms were performed using a custom-built confocal microscope as previously described in ref 8. Briefly, the samples were excited using pulsed-interleaved excitation[33] at wavelengths of 565 and 641 nm (LDH-D-TA-560 and LDH-D-C-640, PicoQuant). We used laser powers of 80 μW for the green laser and 40 μW for the red laser, as measured before the 60× water immersion objective (60×/1.27 WI, Nikon, Düsseldorf, Germany). Fluorescence was filtered by ET607/36 and ET670/30 emission filters (AHF Analysentechnik, Tübingen, Germany) for green and red, respectively.

Measurements were performed in eight-well chamber slides (Nunc Lab-Tek, VWR), which were coated with bovine serum albumin (BSA, New England Biolabs, Ipswich, MA, USA; 20 mg/mL diluted to 1 mg/mL in PBS). The DNA origami samples were diluted to 100−200 pM in imaging buffer (10 mM Tris, 1 mM EDTA, 12 mM $MgCl_2$, as well as 1 mM Trolox[46] and 10 mM cysteamine to reduce photobleaching and blinking). For repair, 600 pM of the DNA origami sample was incubated with 1 μL of T4 DNA ligase (New Englang Biolabs; 400.000 units/mL) in 1× T4 DNA ligase buffer (New England Biolabs; 50 mM Tris, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5) in a total reaction volume of 10 μL at 16 °C overnight or for 1 h at room temperature. For subsequent incubation

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

with the restriction enzyme ZraI, the T4 DNA ligase and EDTA were removed from the buffer with Amicon Ultra centrifugal filters, 100 kDa MWCO (Merck, Darmstadt, Germany), and a buffer composed of 10 mM Tris and 12 mM $MgCl_2$. Around 3.5 nM of the ligated DNA origami samples was incubated with 1 μL of ZraI (New Englang Biolabs, 10.000 units/mL) in 1× CutSmart buffer (New England Biolabs; 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) in a total reaction volume of 20 μL at room temperature overnight. For subsequent incubation with the T4 DNA ligase, ZraI was removed with Amicon Ultra centrifugal filters, 100 kDa MWCO (Merck, Darmstadt, Germany), using a buffer of 10 mM Tris, 1 mM EDTA, and 12 mM $MgCl_2$. For spFRET measurements, samples were diluted to 100−200 pM in imaging buffer.

Data were analyzed using the open-source software PAM (pulsed interleaved excitation analysis with MATLAB).[47] A burst-wise analysis in combination with a two-channel kernel-density estimator[48] (2CDE)-filter allowed us to remove donor- and acceptor-only populations and to focus on double-labeled species. Bursts were selected using the all-photon search algorithm, which is based on a sliding time window approach.[49] Parameters for the burst search were fixed to a minimum of 500 photons per burst, a time window of 2500 μs, and 50 photons per time window. To remove molecules without a donor or an acceptor dye, an ALEX-2CDE filter[48] with a time constant of 100 μs and an upper limit of 15 was used. Accurate FRET efficiencies, E, were calculated by correcting for crosstalk (α = 0.138), direct excitation (δ = 0.125), and differences in the quantum yield and detection efficiency (γ = 0.9) according to standardized procedures.[4]

To calculate error bars for the bar plots in Figures 2D and 5B, we determined the total number of molecules, n, and the fraction of molecules displaying high-FRET efficiency, p (Tables S2 and S4). Assuming a binominal distribution, the uncertainty σ was estimated from the standard deviation according to

$$\sigma = \sqrt{\frac{p(1-p)}{n}} \quad (1)$$

The 95% confidence intervals are given as $\pm 1.96\sigma$.

Single-Pair FRET Measurements Using TIRF Microscopy. Single-pair FRET measurements of surface-immobilized DNA origami platforms were performed on a home-built TIRF microscope, which offers both objective- and prism-type excitation. The setup was recently described in detail for three-color experiments using prism-excitation.[10] Here, we used objective-type TIRF excitation with lasers at 561 nm (Cobolt Jive, 50 mW) and 647 nm (Cobolt MLD, 120 mW), which were alternated on a time scale of 30 ms (+ 3 ms frame transfer). Fluorescence was collected by a 60× oil immersion objective (Apo TIRF 60×/1.49 Oil, Nikon) and separated by a 405/488/561/647 polychroic mirror (AHF Analysentechnik, Tübingen, Germany). Donor and acceptor emission was filtered using HQ 595/50 and HQ 705/100 emission filters (AHF Analysentechnik, Tübingen, Germany), respectively.

The DNA origami platforms were measured in flow chambers, which were assembled using a biotin-PEG silane (Nanocs Inc., NY, USA) functionalized coverslip, a Nesco film spacer, and a microscope slide with two holes drilled through it for in- and outlet tubings. The flow chambers were first incubated with streptavidin (Sigma-Aldrich, St. Louis, MO, USA; 0.2 mg/mL in PBS) for 20 min and washed with origami buffer (10 mM Tris, 1 mM EDTA, and 12 mM $MgCl_2$). The DNA origami samples were diluted to 15 pM in origami buffer and immobilized on the coverslip through biotin−streptavidin−biotin binding. Unbound DNA origamis were washed away with origami buffer. For stabilization of the fluorophores, we flushed in a photococktail containing 10 mM Tris, 1 mM EDTA, 12 mM $MgCl_2$, 1 mM Trolox[46] (UV activated for 5 min), 10% (v/v) glycerol, 1% (w/v) glucose, and 10% (v/v) glucoseoxidase−catalase solution for oxygen removal.[50] For ligation, the surface-immobilized origamis were incubated with roughly 300 nM T4 DNA ligase in 1× T4 DNA ligase buffer (50 mM Tris, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5) for at least 30 min. Afterward, the flow chambers were washed with origami buffer and then flushed with the photococktail. For the live ligation experiments, the T4 DNA ligase was diluted to 600 nM in 1× T4 DNA ligase buffer and subsequently diluted to roughly 100 pM in buffer containing the photococktail. For the ligation and subsequent digestion experiments, custom-built flow chambers were fixed to the microscope with adhesive tape to ensure no movement of the sample on the microscope during washing with different enzyme-containing buffers. Thus, the same field of view could be observed for the duration of the experiment. After the 30 min incubation step with 300 nM T4 DNA ligase as described above, the flow chambers were washed with a buffer of 10 mM Tris and 12 mM $MgCl_2$. The origamis were then incubated with 2 μL of ZraI (New Englang Biolabs, 10.000 units/mL) in 1× CutSmart buffer (New England Biolabs; 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) for 1 h at room temperature. After a washing step with origami buffer, the photococktail was flushed in for imaging. After an additional washing step with origami buffer, a second incubation step with 300 nM T4 DNA ligase followed, this time for 1 h. After a washing step with origami buffer, the photococktail was flushed in for imaging.

Data were recorded with an alternating excitation sequence of green−red with typical laser powers of 12.5 mW for the green 561 nm laser and 20 mW for the red 647 nm laser at an exposure time of 30 ms for 4000 frames. For the ligation and digestion experiment, only 500 frames were recorded per condition to reduce photobleaching. TIRF data were analyzed with custom-written MATLAB programs (Mathworks, MA, USA). For analysis we exclusively chose double-labeled molecules showing fluorescence intensity after both donor and acceptor excitation and excluded acceptor- or donor-only species. For data correction, we typically used α = 0.138 to account for spectral cross-talk and γ = 0.6 for differences in relative detection efficiencies.

AFM Imaging. AFM images were recorded in solution on a NanoWizard 3 Ultra AFM (JPK Instruments AG, Berlin, Germany) in TAE buffer containing 40 mM Tris, 2 mM EDTA, and 12.5 mM $MgCl_2$. A mica surface (Quality V1, Plano GmbH, Wetzlar, Germany) was freshly cleaved, incubated with a 10 mM $NiCl_2$ solution for 5 min, rinsed with double-distilled water, and dried with air. Next, 10 μL of a 1 nM DNA origami solution was added to the surface and incubated for 3 min. After three washing steps with TAE buffer, data were recorded in tapping mode using an ultrashort cantilever, USC-F0.3-k0.3 (Nanoworld via NanoAndMore GmbH, Wetzlar, Germany), at a resolution of 512 × 512 pixels. AFM images were analyzed with a JPK data processing software (JPK Instruments AG, Berlin, Germany).

Coarse-Grained Molecular Dynamics Simulations. Coarse-grained structural models of the DNA origami frame were generated from the caDNAno[51] JSON files using scripts provided with the oxDNA package.[31] The initial structural model exhibited abnormally stretched bonds. To relax the structure, we followed the procedure described in ref 52. Stretched bonds were broken and replaced by harmonic potential restraints of increasing strength during relaxation. After relaxation of the origami structure, hybridization of the single-stranded DNA overhangs was induced by applying weak harmonic potential to the complementary bases, resulting in the unligated state (Figure S2A). Molecular dynamics simulations were performed using the oxDNA package[31] using the oxDNA2 model[53] at a temperature of 300 K and a monovalent salt concentration of 500 mM Na^+ to mimic the electrostatic screening of the experimental Mg^{2+} concentrations of 12 mM. Simulations were performed using a time step of 15.15 fs using an Andersen-like thermostat every 103 Newtonian time steps. For the unligated and ligated states, we performed ∼6 μs long simulations. Due to the coarse-grained model chosen and the implicit solvent treatment, the dynamics are sped up in the simulation. A conversion factor between simulated and physical time scales may be obtained by comparing the diffusion coefficient obtained from the simulations to experimental values,[54] resulting in an approximative speedup by a factor of ∼330 under the conditions of the simulation.[52] Thus, the simulations represent a time scale of ∼2 ms with a time step of 5 ps.

Accessible Volume Simulations. A double-stranded B-DNA model was generated using the Nucleic Acid Builder (NAB) program

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

for Amber.[55] Next, the labeling positions on the C7 atoms of the respective thymine bases were identified. The accessible volumes[35] for Atto647N acceptor and Cy3b donor dyes were then calculated with the FRET positioning and screening (FPS) software.[36] We chose the three radii AV3 model and the dye parameters summarized in Table S6, based on recommendations from ref 4. To compute FRET-averaged distances and expected FRET efficiencies, we assumed a Förster radius of 67 Å for the Cy3b/Atto647N FRET pair.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.9b09093.

Additional AFM image, coarse-grained MD simulations, characterization of freely diffusing DNA origami platforms, additional spFRET efficiency histograms, dwell time histograms, additional FRET efficiency traces before and after ligation, dye accessible volumes on dsDNA, schematic of dsDNA for repair, TIRFM fields of view, additional FRET efficiency traces after ligation and subsequent cleavage, DNA sequences, burst statistics, and parameters for accessible volume simulations (PDF)

Coarse-grained MD movie of the DNA origami (MOV)

Coarse-grained MD movie of the DNA origami (MOV)

■ AUTHOR INFORMATION

Corresponding Author

*d.lamb@lmu.de

ORCID

Anders Barth: 0000-0003-3671-3072

Tim Liedl: 0000-0002-0040-0173

Don C. Lamb: 0000-0002-0232-1903

Present Addresses

[§]Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[⊥]Department für Physik, Université de Fribourg, 1700 Fribourg, Switzerland.

[∥]European Molecular Biology Laboratory (EMBL) Rome, 00015 Monterotondo, Italy.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank J. Bohlen, D. Rüdiger, and A. Kostina for AFM images. We are grateful to E. Plötz for sharing her flow chamber preparation protocol for TIRFM and to B. Salem for providing analysis software. K.B, A.B. and D.C.L. wish to thank Dr. Klaus Römer for his support over the years. We further thank R. Jungmann and K. Duderstadt for fruitful discussions. K.B. acknowledges support by the IMPRS-LS graduate school. Financial support was provided from the Deutsche Forschungsgemeinschaft (DFG) through the SFB1032 (Project A6) and GRK1721 (project C3) and the Ludwig-Maximilians-Universität München via the Center for NanoScience Munich (CeNS) and LMUinnovativ initiative BioImaging Network (BIN).

■ REFERENCES

(1) Förster, T. Zwischenmolekulare Energiewanderung Und Fluoreszenz. *Ann. Phys. (Berlin, Ger.)* **1948**, *2*, 55−75.

(2) Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S. Toward Dynamic Structural Biology: Two Decades of Single-Molecule Förster Resonance Energy Transfer. *Science* **2018**, *359*, No. eaan1133.

(3) Ha, T.; Enderle, T.; Ogletree, D. F.; Chemla, D. S.; Selvin, P. R.; Weiss, S. Probing the Interaction Between Two Single Molecules: Fluorescence Resonance Energy Transfer Between a Single Donor and a Single Acceptor. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6264−6268.

(4) Hellenkamp, B.; Schmid, S.; Doroshenko, O.; Opanasyuk, O.; Kühnemuth, R.; Rezaei Adariani, S.; Ambrose, B.; Aznauryan, M.; Barth, A.; Birkedal, V.; Bowen, M. E.; Chen, H.; Cordes, T.; Eilert, T.; Fijen, C.; Gebhardt, C.; Götz, M.; Gouridis, G.; Gratton, E.; Ha, T.; Hao, P.; Hanke, C. A.; Hartmann, A.; Hendrix, J.; Hildebrandt, L. L.; Hirschfeld, V.; Hohlbein, J.; Hua, B.; Hübner, C. G.; Kallis, E.; Kapanidis, A. N.; Kim, J.-Y.; Krainer, G.; Lamb, D. C.; Lee, N. K.; Lemke, E. A.; Levesque, B.; Levitus, M.; McCann, J. J.; Naredi-Rainer, N.; Nettels, D.; Ngo, T.; Qiu, R.; Robb, N. C.; Röcker, C.; Sanabria, H.; Schlierf, M.; Schröder, T.; Schuler, B.; Seidel, H.; Streit, L.; Thurn, J.; Tinnefeld, P.; Tyagi, S.; Vandenberk, N.; Vera, A. M.; Weninger, K. R.; Wünsch, B.; Yanez-Orozco, I. S.; Michaelis, J.; Seidel, C. A. M.; Craggs, T. D.; Hugel, T. Precision and Accuracy of Single-Molecule FRET Measurements-a Multi-Laboratory Benchmark Study. *Nat. Methods* **2018**, *15*, 669−676.

(5) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297−302.

(6) Stein, I. H.; Schüller, V.; Böhm, P.; Tinnefeld, P.; Liedl, T. Single-Molecule FRET Ruler Based on Rigid DNA Origami Blocks. *ChemPhysChem* **2011**, *12*, 689−695.

(7) Pibiri, E.; Holzmeister, P.; Lalkens, B.; Acuna, G. P.; Tinnefeld, P. Single-Molecule Positioning in Zeromode Waveguides by DNA Origami Nanoadapters. *Nano Lett.* **2014**, *14*, 3499−3503.

(8) Nicoli, F.; Barth, A.; Bae, W.; Neukirchinger, F.; Crevenna, A. H.; Lamb, D. C.; Liedl, T. Directional Photonic Wire Mediated by Homo-Förster Resonance Energy Transfer on a DNA Origami Platform. *ACS Nano* **2017**, *11*, 11264−11272.

(9) Saccà, B.; Meyer, R.; Erkelenz, M.; Kiko, K.; Arndt, A.; Schroeder, H.; Rabe, K. S.; Niemeyer, C. M. Orthogonal Protein Decoration of DNA Origami. *Angew. Chem., Int. Ed.* **2010**, *49*, 9378−9383.

(10) Kopperger, E.; List, J.; Madhira, S.; Rothfischer, F.; Lamb, D. C.; Simmel, F. C. A Self-Assembled Nanoscale Robotic Arm Controlled by Electric Fields. *Science* **2018**, *359*, 296−301.

(11) Funke, J. J.; Dietz, H. Placing Molecules with Bohr Radius Resolution Using DNA Origami. *Nat. Nanotechnol.* **2016**, *11*, 47−52.

(12) Hartl, C.; Frank, K.; Amenitsch, H.; Fischer, S.; Liedl, T.; Nickel, B. Position Accuracy of Gold Nanoparticles on DNA Origami Structures Studied with Small-Angle X-Ray Scattering. *Nano Lett.* **2018**, *18*, 2609−2615.

(13) Sannohe, Y.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. Visualization of Dynamic Conformational Switching of the G-Quadruplex in a DNA Nanostructure. *J. Am. Chem. Soc.* **2010**, *132*, 16311−16313.

(14) Selnihhin, D.; Sparvath, S. M.; Preus, S.; Birkedal, V.; Andersen, E. S. Multifluorophore DNA Origami Beacon as a Biosensing Platform. *ACS Nano* **2018**, *12*, 5699−5708.

(15) Khara, D. C.; Schreck, J. S.; Tomov, T. E.; Berger, Y.; Ouldridge, T. E.; Doye, J. P. K.; Nir, E. DNA Bipedal Motor Walking Dynamics: an Experimental and Theoretical Study of the Dependency on Step Size. *Nucleic Acids Res.* **2018**, *46*, 1553−1561.

(16) Nickels, P. C.; Wünsch, B.; Holzmeister, P.; Bae, W.; Kneer, L. M.; Grohmann, D.; Tinnefeld, P.; Liedl, T. Molecular Force Spectroscopy with a DNA Origami-Based Nanoscopic Force Clamp. *Science* **2016**, *354*, 305−307.

(17) Aparicio, T.; Baer, R.; Gautier, J. DNA Double-Strand Break Repair Pathway Choice and Cancer. *DNA Repair* **2014**, *19*, 169−175.

(18) Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annu. Rev. Biochem.* **2010**, *79*, 181−211.

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

(19) Pastwa, E.; Somiari, R. I.; Malinowski, M.; Somiari, S. B.; Winters, T. A. In Vitro Non-Homologous DNA End Joining Assays–the 20th Anniversary. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 1254−1260.
(20) Mari, P.-O.; Florea, B. I.; Persengiev, S. P.; Verkaik, N. S.; Brüggenwirth, H. T.; Modesti, M.; Giglia-Mari, G.; Bezstarosti, K.; Demmers, J. A. A.; Luider, T. M.; Houtsmuller, A. B.; van Gent, D. C. Dynamic Assembly of End-Joining Complexes Requires Interaction Between Ku70/80 and XRCC4. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18597−18602.
(21) Yano, K.-I.; Chen, D. J. Live Cell Imaging of XLF and XRCC4 Reveals a Novel View of Protein Assembly in the Non-Homologous End-Joining Pathway. *Cell Cycle* **2008**, *7*, 1321−1325.
(22) Britton, S.; Coates, J.; Jackson, S. P. A New Method for High-Resolution Imaging of Ku Foci to Decipher Mechanisms of DNA Double-Strand Break Repair. *J. Cell Biol.* **2013**, *202*, 579−595.
(23) Abdisalaam, S.; Davis, A. J.; Chen, D. J.; Alexandrakis, G. Scanning Fluorescence Correlation Spectroscopy Techniques to Quantify the Kinetics of DNA Double Strand Break Repair Proteins After Γ-Irradiation and Bleomycin Treatment. *Nucleic Acids Res.* **2014**, *42*, No. e5-e5.
(24) Reid, D. A.; Keegan, S.; Leo-Macias, A.; Watanabe, G.; Strande, N. T.; Chang, H. H.; Oksuz, B. A.; Fenyo, D.; Lieber, M. R.; Ramsden, D. A.; Rothenberg, E. Organization and Dynamics of the Nonhomologous End-Joining Machinery During DNA Double-Strand Break Repair. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2575−E2584.
(25) Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. A Versatile DNA Nanochip for Direct Analysis of DNA Base-Excision Repair. *Angew. Chem., Int. Ed.* **2010**, *49*, 9412−9416.
(26) Endo, M.; Takeuchi, Y.; Suzuki, Y.; Emura, T.; Hidaka, K.; Wang, F.; Willner, I.; Sugiyama, H. Single-Molecule Visualization of the Activity of a Zn2+-Dependent DNAzyme. *Angew. Chem., Int. Ed.* **2015**, *54*, 10550−10554.
(27) Takeuchi, Y.; Endo, M.; Suzuki, Y.; Hidaka, K.; Durand, G.; Dausse, E.; Toulmé, J.-J.; Sugiyama, H. Single-Molecule Observations of RNA-RNA Kissing Interactions in a DNA Nanostructure. *Biomater. Sci.* **2016**, *4*, 130−135.
(28) Shuman, S. DNA Ligases: Progress and Prospects. *J. Biol. Chem.* **2009**, *284*, 17365−17369.
(29) Lehman, I. R. DNA Ligase: Structure, Mechanism, and Function. *Science* **1974**, *186*, 790−797.
(30) Povirk, L. F. Processing of Damaged DNA Ends for Double-Strand Break Repair in Mammalian Cells. *ISRN Mol. Biol.* **2012**, *2012*.
(31) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model. *J. Chem. Phys.* **2011**, *134*, 085101.
(32) Müller, B. K.; Zaychikov, E.; Bräuchle, C.; Lamb, D. C. Pulsed Interleaved Excitation. *Biophys. J.* **2005**, *89*, 3508−3522.
(33) Kudryavtsev, V.; Sikor, M.; Kalinin, S.; Mokranjac, D.; Seidel, C. A. M.; Lamb, D. C. Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem* **2012**, *13*, 1060−1078.
(34) Antonik, M.; Felekyan, S.; Gaiduk, A.; Seidel, C. A. M. Separating Structural Heterogeneities From Stochastic Variations in Fluorescence Resonance Energy Transfer Distributions via Photon Distribution Analysis. *J. Phys. Chem. B* **2006**, *110*, 6970−6978.
(35) Sindbert, S.; Kalinin, S.; Nguyen, H.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Müller, S.; Seidel, C. A. M. Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.* **2011**, *133*, 2463−2480.
(36) Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A Toolkit and Benchmark Study for FRET-Restrained High-Precision Structural Modeling. *Nat. Methods* **2012**, *9*, 1218−1225.
(37) Ramakrishnan, S.; Schärfen, L.; Hunold, K.; Fricke, S.; Grundmeier, G.; Schlierf, M.; Keller, A.; Krainer, G. Enhancing the Stability of DNA Origami Nanostructures: Staple Strand Redesign Versus Enzymatic Ligation. *Nanoscale* **2019**, *11*, 16270−16276.
(38) Schickinger, M.; Zacharias, M.; Dietz, H. Tethered Multi-fluorophore Motion Reveals Equilibrium Transition Kinetics of Single DNA Double Helices. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E7512−E7521.
(39) Potapov, V.; Ong, J. L.; Langhorst, B. W.; Bilotti, K.; Cahoon, D.; Canton, B.; Knight, T. F.; Evans, T. C.; Lohman, G. J. S. A Single-Molecule Sequencing Assay for the Comprehensive Profiling of T4 DNA Ligase Fidelity and Bias During DNA End-Joining. *Nucleic Acids Res.* **2018**, *46*, No. e79-e79.
(40) Kapanidis, A. N.; Laurence, T. A.; Lee, N. K.; Margeat, E.; Kong, X.; Weiss, S. Alternating-Laser Excitation of Single Molecules. *Acc. Chem. Res.* **2005**, *38*, 523−533.
(41) Rabiner, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* **1989**, *77*, 257−286.
(42) McKinney, S. A.; Joo, C.; Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys. J.* **2006**, *91*, 1941−1951.
(43) Zarrabi, N.; Schluesche, P.; Meisterernst, M.; Börsch, M.; Lamb, D. C. Analyzing the Dynamics of Single TBP-DNA-NC2 Complexes Using Hidden Markov Models. *Biophys. J.* **2018**, *115*, 2310−2326.
(44) Trads, J. B.; Tørring, T.; Gothelf, K. V. Site-Selective Conjugation of Native Proteins with DNA. *Acc. Chem. Res.* **2017**, *50*, 1367−1374.
(45) Stahl, E.; Martin, T. G.; Praetorius, F.; Dietz, H. Facile and Scalable Preparation of Pure and Dense DNA Origami Solutions. *Angew. Chem., Int. Ed.* **2014**, *53*, 12735−12740.
(46) Cordes, T.; Vogelsang, J.; Tinnefeld, P. On the Mechanism of Trolox as Antiblinking and Antibleaching Reagent. *J. Am. Chem. Soc.* **2009**, *131*, 5018−5019.
(47) Schrimpf, W.; Barth, A.; Hendrix, J.; Lamb, D. C. PAM: a Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data. *Biophys. J.* **2018**, *114*, 1518−1528.
(48) Tomov, T. E.; Tsukanov, R.; Masoud, R.; Liber, M.; Plavner, N.; Nir, E. Disentangling Subpopulations in Single-Molecule FRET and ALEX Experiments with Photon Distribution Analysis. *Biophys. J.* **2012**, *102*, 1163−1173.
(49) Nir, E.; Michalet, X.; Hamadani, K. M.; Laurence, T. A.; Neuhauser, D.; Kovchegov, Y.; Weiss, S. Shot-Noise Limited Single-Molecule FRET Histograms: Comparison Between Theory and Experiments. *J. Phys. Chem. B* **2006**, *110*, 22103−22124.
(50) Stein, I. H.; Capone, S.; Smit, J. H.; Baumann, F.; Cordes, T.; Tinnefeld, P. Linking Single-Molecule Blinking to Chromophore Structure and Redox Potentials. *ChemPhysChem* **2012**, *13*, 931−937.
(51) Douglas, S. M.; Marblestone, A. H.; Teerapittayanon, S.; Vazquez, A.; Church, G. M.; Shih, W. M. Rapid Prototyping of 3D DNA-Origami Shapes with caDNAno. *Nucleic Acids Res.* **2009**, *37*, 5001−5006.
(52) Shi, Z.; Castro, C. E.; Arya, G. Conformational Dynamics of Mechanically Compliant DNA Nanostructures From Coarse-Grained Molecular Dynamics Simulations. *ACS Nano* **2017**, *11*, 4617−4630.
(53) Snodin, B. E. K.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J. S.; Romano, F.; Ouldridge, T. E.; Tsukanov, R.; Nir, E.; Louis, A. A.; Doye, J. P. K. Introducing Improved Structural Properties and Salt Dependence Into a Coarse-Grained Model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901.
(54) Agrawal, V.; Arya, G.; Oswald, J. Simultaneous Iterative Boltzmann Inversion for Coarse-Graining of Polyurea. *Macromolecules* **2014**, *47*, 3378−3389.
(55) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: a Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13*, 55−58.

DOI: 10.1021/jacs.9b09093
J. Am. Chem. Soc. 2020, 142, 815−825

Supporting Information

A DNA origami platform for single-pair Förster Resonance Energy Transfer investigation of DNA-DNA interactions and ligation

Kira Bartnik,[†] Anders Barth,[†,§] Mauricio Pilo-Pais,[‡,⊥] Alvaro H. Crevenna,[†,∥] Tim Liedl,[‡] and Don C. Lamb[*,†]

[†]Department of Chemistry, Center for Nanoscience (CeNS), Nanosystems Initiative Munich (NIM) and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany

[‡]Department of Physics and Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität, 80539 Munich, Germany

Present Addresses

[§]Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany

[⊥]Department für Physik, Université de Fribourg, 1700 Fribourg, Switzerland

[∥] European Molecular Biology Laboratory (EMBL) Rome, 00015 Monterotondo, Italy

* Corresponding author, E-Mail: d.lamb@lmu.de

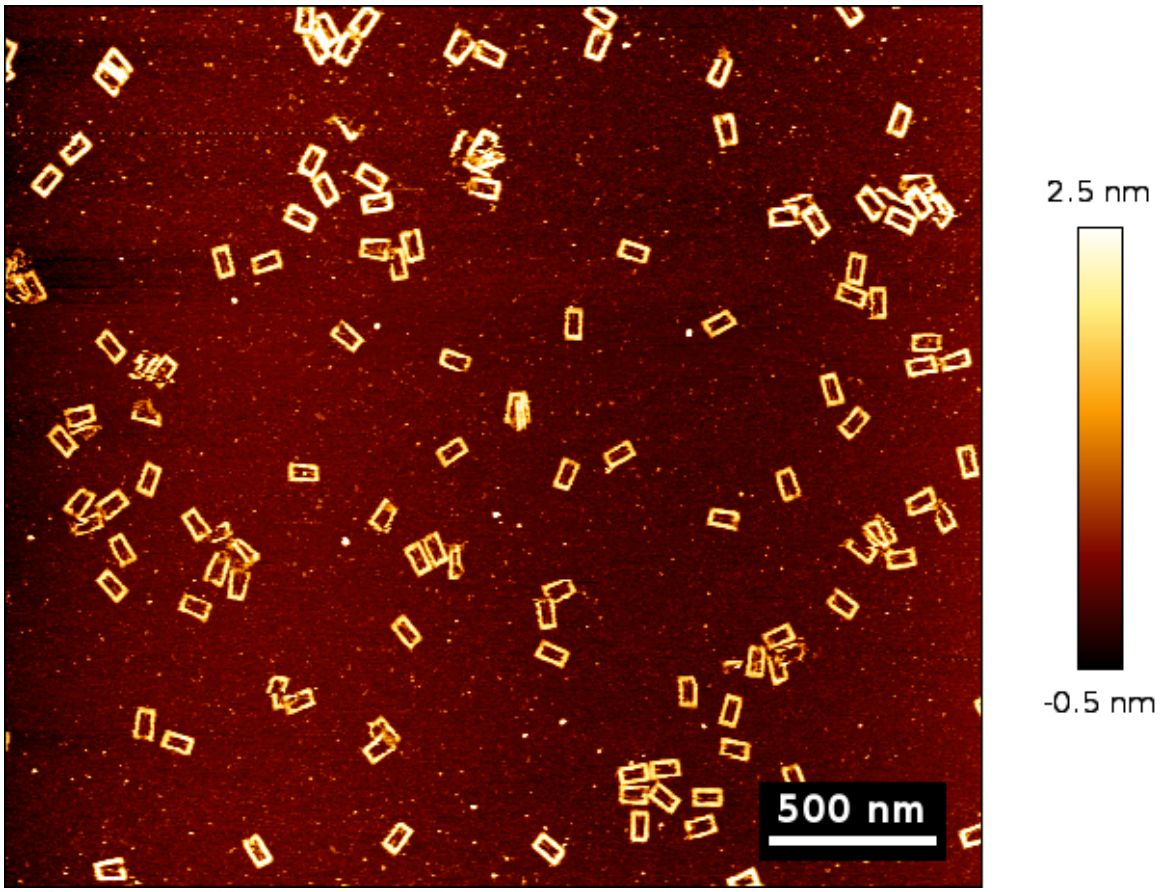


Figure S1. An AFM image of the folded DNA origami frames immobilized on a mica surface. The DNA origami platforms shown were incubated with the T4 DNA ligase before immobilization.

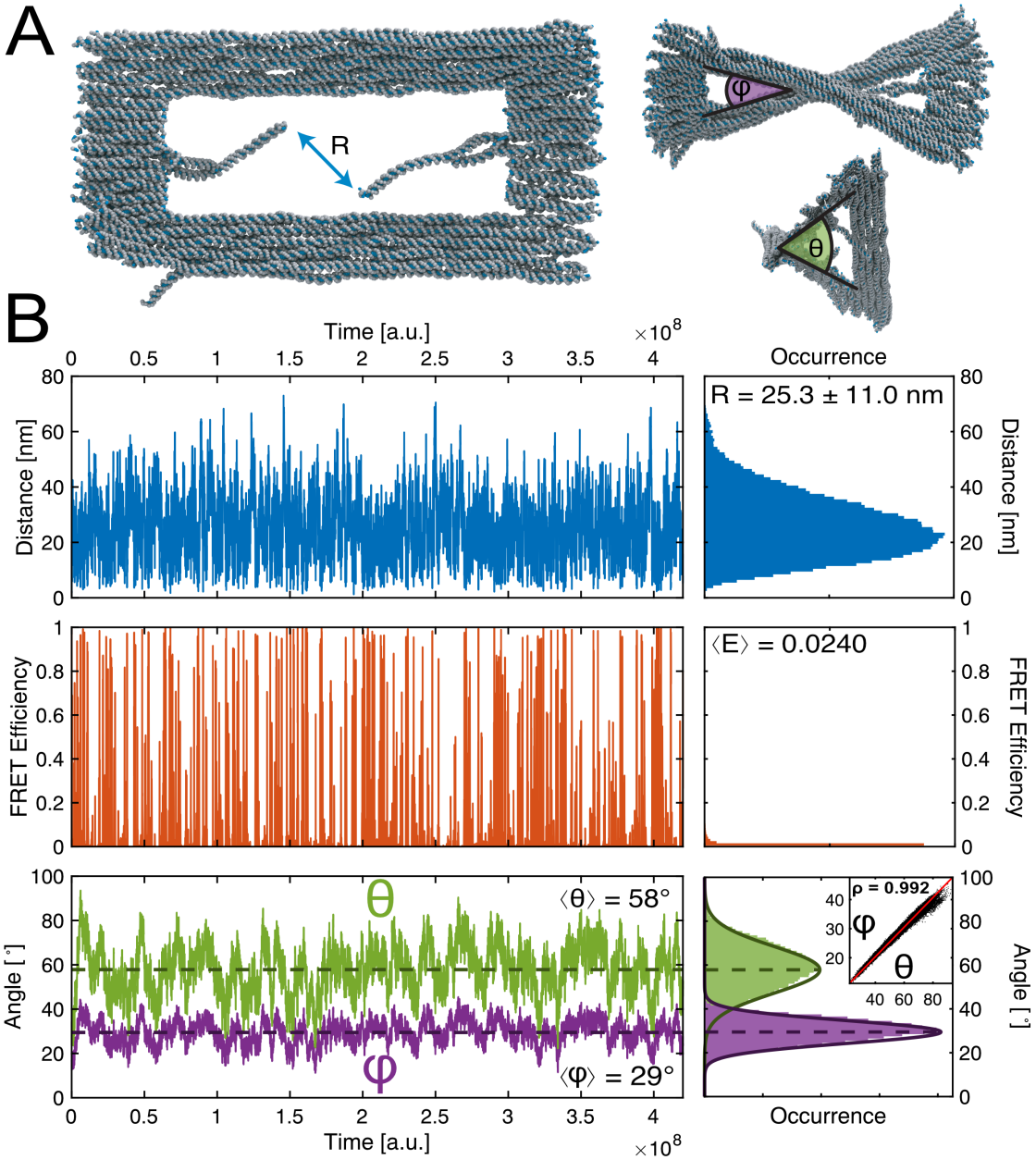


Figure S2. End-to-end distance distributions and twist angles of the origami frame with 4-nt overhangs obtained from coarse-grained MD simulations. **A**) Shown is a snapshot of the MD simulation at time step $t = 2 \times 10^7$. The end-to-end distance R between the overhangs is calculated from the terminal residues of each strand (blue arrow). During the simulation, the origami frame underwent a twisting motion, which is monitored by the two angles φ and θ defining the twist between the longer sides and the shorter sides of the frame, respectively. **B**) The time traces of the end-to-end distance (top) and the corresponding FRET efficiency (middle, $R_0 = 67$ Å) reveal large distance fluctuations with an average distance of $R = 25.3 \pm 11.0$ nm. Due to the large separation, the FRET efficiency is found close to zero ($\langle E \rangle = 0.0240$). While the ends are occasionally found in close contact, as indicated by short periods of high FRET efficiency, no stable binding was observed during the simulation. The angles φ and θ reveal large scale structural fluctuations of the origami structure through a twisting motion (bottom). The corresponding distributions are shown to the right. For the angles φ and θ, the distributions are additionally fitted to normal distributions, resulting in average angles of $\langle \varphi \rangle = 58 \pm 11°$ and $\langle \theta \rangle = 29 \pm 5°$. The angles φ and θ are highly correlated (see inset, Pearson correlation coefficient $\rho = 0.992$), indicating that the twisting motion globally affects the whole origami structure.

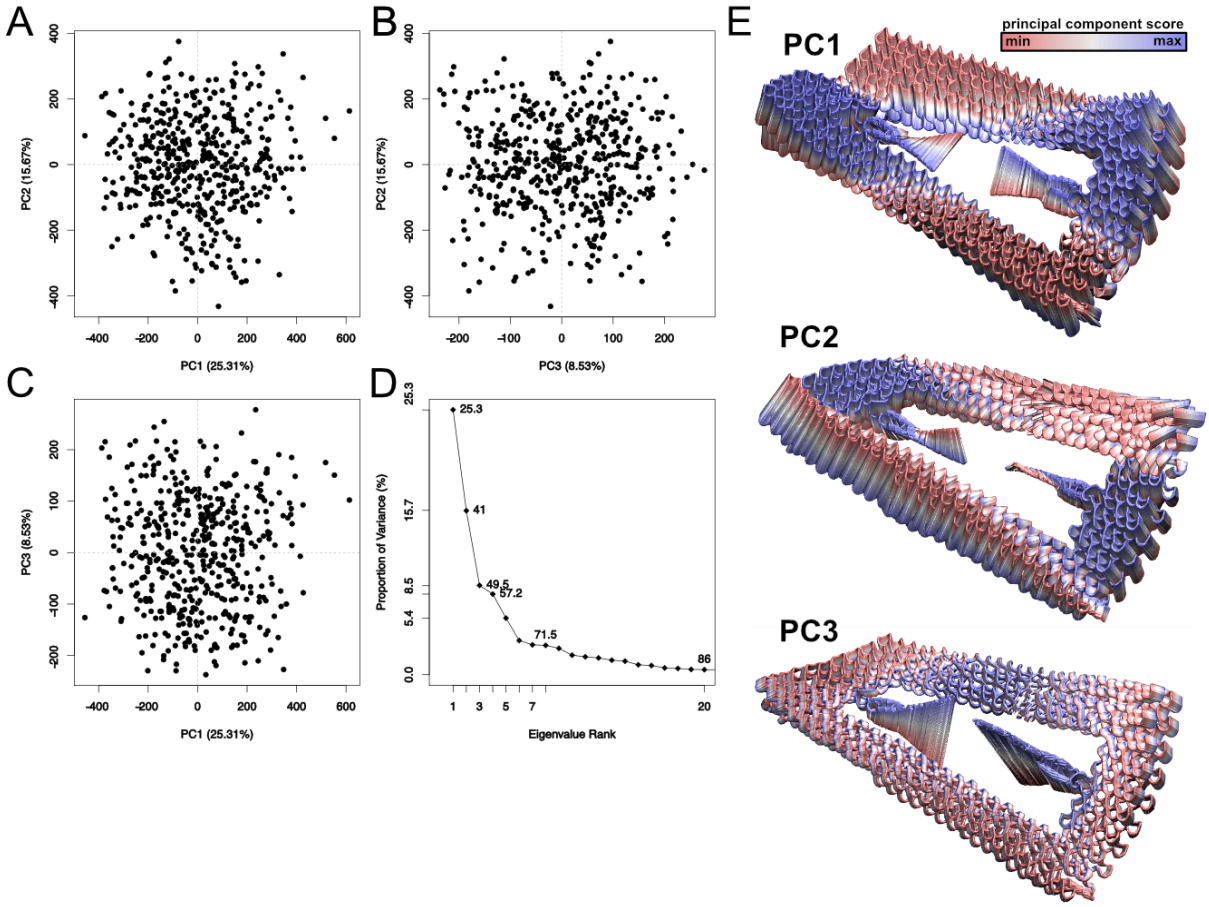


Figure S3. Principle component analysis (PCA) of the coarse-grained MD simulations reveals the predominant global motions of the structure. PCA provides a lower-dimensional description of the conformational dynamics by decomposing the complex global motions into distinct modes that are ranked by the proportion of the structural variance that they account for. **A-C**) Correlation plots of the first three principal components (PC) identified from the trajectory. No correlation is observed between the different components, indicating that they describe independent modes of movement of the origami structure. **D**) Proportion of the explained variance of the different principal components. Approximately 50 % of the structural variance is explained by the first three principal components. **E**) Structural variation of the origami structure along the first three principal components. Snapshots of the motion along the modes are superimposed and colored according to their score along the PC, as indicated by the color bar. PC1 and PC2 contain mainly contributions of the movement of the origami frame and describe two different twisting modes. PC3 and beyond mainly describe the motion of the DNA overhangs in the center of the origami frame. The motions are amplified by a factor of three for illustration purposes. Principle component analysis was performed using the *bio3d* package for R[1] based on 500 structures extracted from the trajectory at evenly-spaced time intervals.

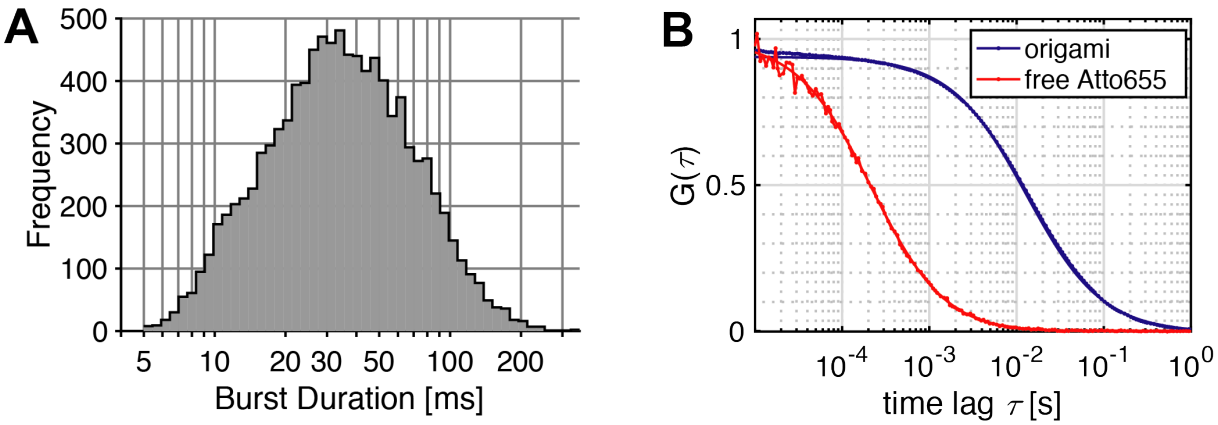


Figure S4. Characterization of freely diffusing DNA origami platforms in solution. **A**) Histogram of fluorescence burst duration with an average duration of around 35 ms. **B**) Typical autocorrelation functions (ACF) of the red detection are shown for an Atto655 dye (diffusion coefficient D = 373 $\mu m^2/s$, shown in red) and a DNA origami with Cy3b- and Atto647N-labeled dsDNAs (D = 5 $\mu m^2/s$, shown in blue). The ACF of the origami is shifted to the right due to the slower diffusion. Measurements were performed using origamis with bound blunt-end DNA double strands.

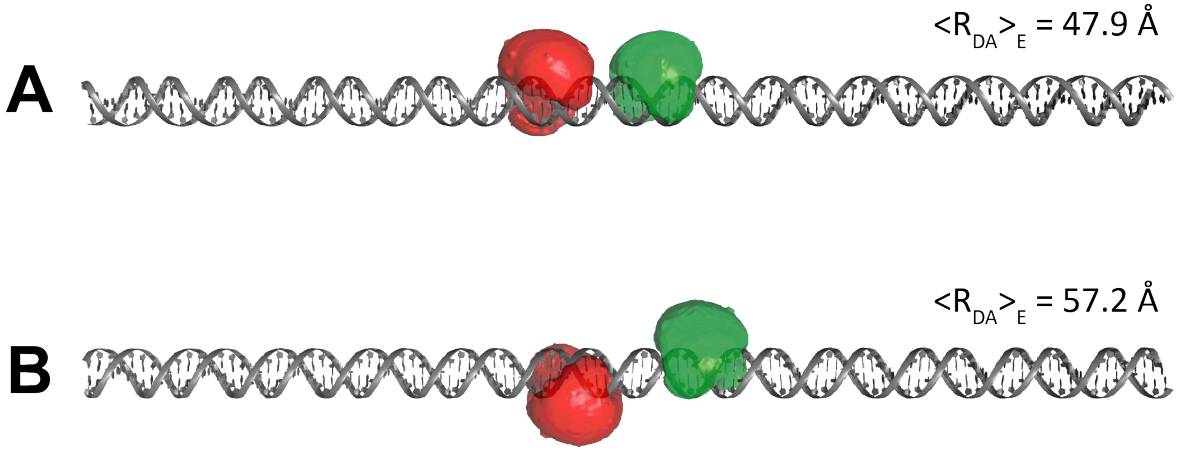


Figure S5. B-DNA model with dye accessible volumes of the acceptor (Atto647N, red) and donor (Cy3b, green) on ligated dsDNA. **A**) Dye accessible volumes on the regular dsDNA sequence are located on the same side of the double-helix. The expected inter-dye distance is 47.9 Å and we determined a distance of 52.9 Å from our experimental data. **B**) Dye accessible volumes on the dsDNA sequence used for the restriction enzyme experiment are on opposite sides of the double-helix. The expected inter-dye distance is 57.2 Å and we determined a distance of 59.7 Å from our experimental data.

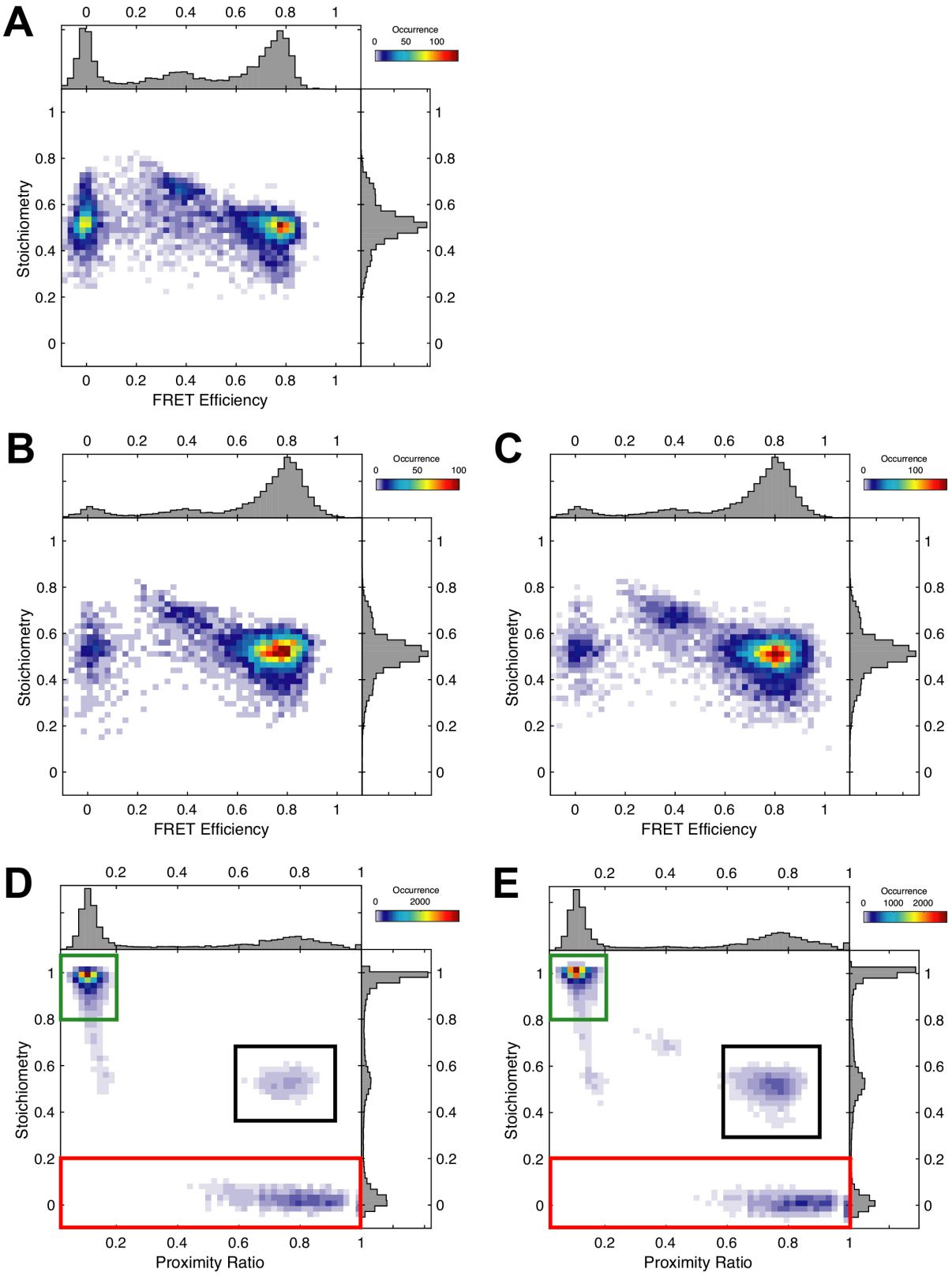


Figure S6. 2D histograms of stoichiometry versus FRET efficiency for DNA origamis with bound 4-nt overhang dsDNA (**A**) or isolated 4-nt control dsDNAs (**B** and **C**) after incubation with the T4 DNA ligase at 16° C over night. 3889 out of 8161 selected molecules in A showed a high-FRET efficiency with E > 0.6, resulting in a high-FRET fraction of 47.65%. While the same DNA concentration (600 pM) and the same T4 DNA ligase concentration according to our standard protocol were used for the ligation experiments reported in A and B, the DNA

concentration in C was increased to 20 nM and a high concentrated T4 DNA ligase (New England Biolabs, Ipswich, Massachusetts, USA; 2.000.000 units/mL instead of 400.000 units/mL) was used, resulting in a 15-fold increase in final ligase concentration compared to our standard ligation protocol. Double-labeled species were selected using a 2CDE-filter ($ALEX_{2CDE} < 15$ for A and $ALEX_{2CDE} < 10$ for B and C). **D**) and **E**): Same data sets as in B and C, respectively, but not filtered for double-labeled species. Hence, donor-only and acceptor-only populations (highlighted by green and red boxes, respectively) are detected. Instead of the FRET efficiency, the proximity ratio (i.e. the uncorrected FRET efficiency) is plotted, as the corrected FRET efficiency is not defined for acceptor-only species. To quantify the fraction of high-FRET efficiency events (highlighted by a black box), used as a measure for the apparent ligation efficiency, we employed a slightly different approach compared to double strands bound to the origami. As the free DNA double strands do not co-diffuse before ligation, the number of double-labeled molecules does not equal the maximum number of molecules that can potentially be ligated. Thus, we have to define this maximum number differently. To this end, we determined the numbers of donor-only and acceptor-only molecules. Due to slightly different concentrations of donor- and acceptor-labeled DNA double strands, these numbers were not identical. Thus, we took the lower number (in our control experiments it was the number of the acceptor-only events) and assumed it to be the number of unligated strands. The apparent ligation efficiency was then calculated by dividing the number of high-FRET efficiency events by the sum of high-FRET efficiency and acceptor-only events. The numbers and the normalized high-FRET fractions for the free DNA double strands are given in Table S3. Under the same conditions (identical DNA and T4 DNA ligase concentrations), we observed significantly less ligation for the free DNA double strands. The high FRET fraction was 48% versus 29% for the DNA origami and free DNA, respectively. When increasing the concentration of the isolated control dsDNAs by a factor of 30, we obtained similar apparent FRET efficiencies with or without the origami platform (48% and 45%, respectively). This indicates that not only does our platform not interfere with the ligation efficiency, but that we achieved our goal of increasing the location concentration of DNA and thus the ability to investigate DNA ligation using a DNA origami platform.

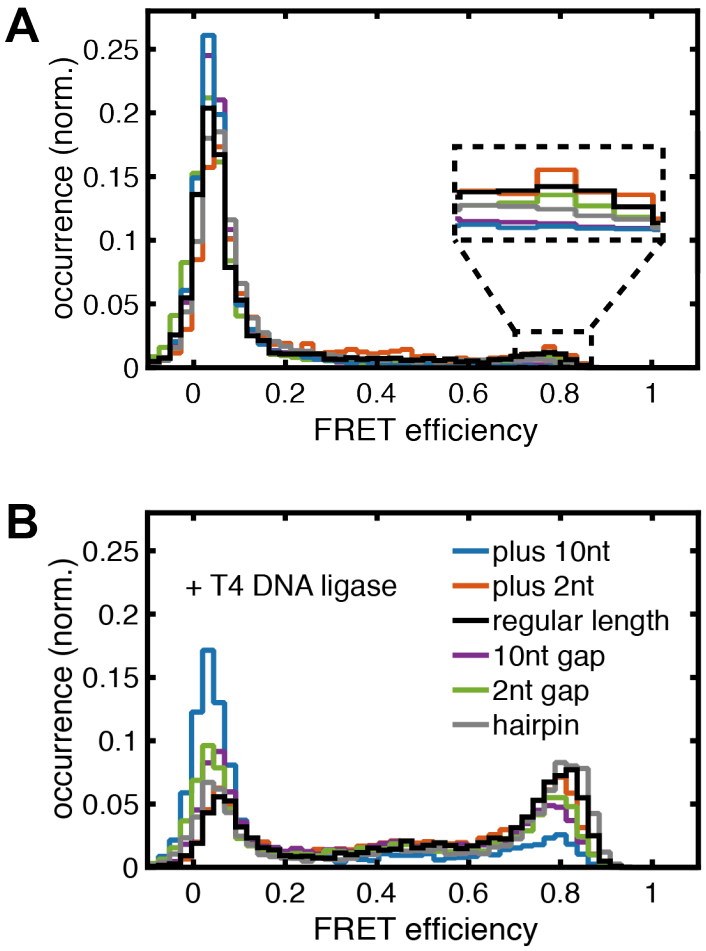


Figure S7. SpFRET efficiency histograms of freely diffusing DNA origami platforms with bound DNA double strands of various lengths before (**A**) and after incubation with the T4 DNA ligase (**B**). The length of the original dsDNA strand was chosen such that, upon ligation, it would exactly span the gap within the DNA origami frame (referred to as "regular length"). Here, we also tested DNA strands that were 2 or 10 nt longer ("plus 2" or "plus 10 nt", respectively), 2 or 10 nt shorter than the regular length dsDNA ("2 nt gap" or "10 nt gap", respectively) or with an additional hairpin sequence ("hairpin"). All DNA sequences had complementary overhangs of 4 nt. An overview of the DNA sequences can be found in table S1. **A**) Before ligation, dsDNAs which were either 10 nt longer or shorter than the original design show no high-FRET population and therefore no (or very short) transient binding of the complementary overhangs. Elongating or shortening of the DNA sequence by 2 nt or the addition of a hairpin sequence has no striking effect compared to the regular length. **B**) All dsDNA lengths show a shift from the low-FRET to the high-FRET state after ligation with the T4 DNA ligase. Ligation is most efficient for regular length dsDNAs and the hairpin sequence. Only the plus 10 nt dsDNA shows a significant decrease in ligation efficiency.

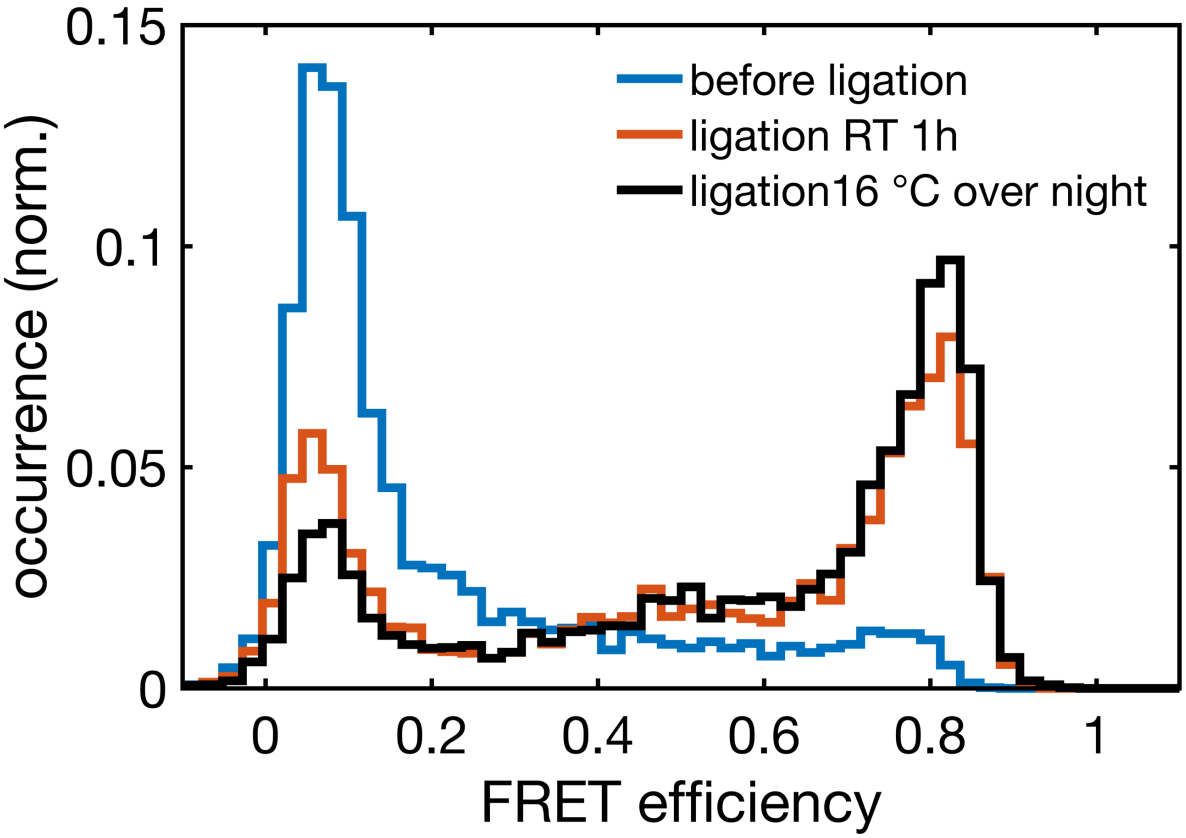


Figure S8. SpFRET efficiency histograms of freely diffusing DNA origami platforms with bound dsDNAs before and after incubation with the T4 DNA ligase. Incubation with the T4 DNA ligase at 16° C overnight (black line) leads to a more pronounced high-FRET population compared to an incubation at room temperature for one hour (orange line) and therefore to more ligated molecules. Complementary overhangs of 4 nt were chosen for this experiment.

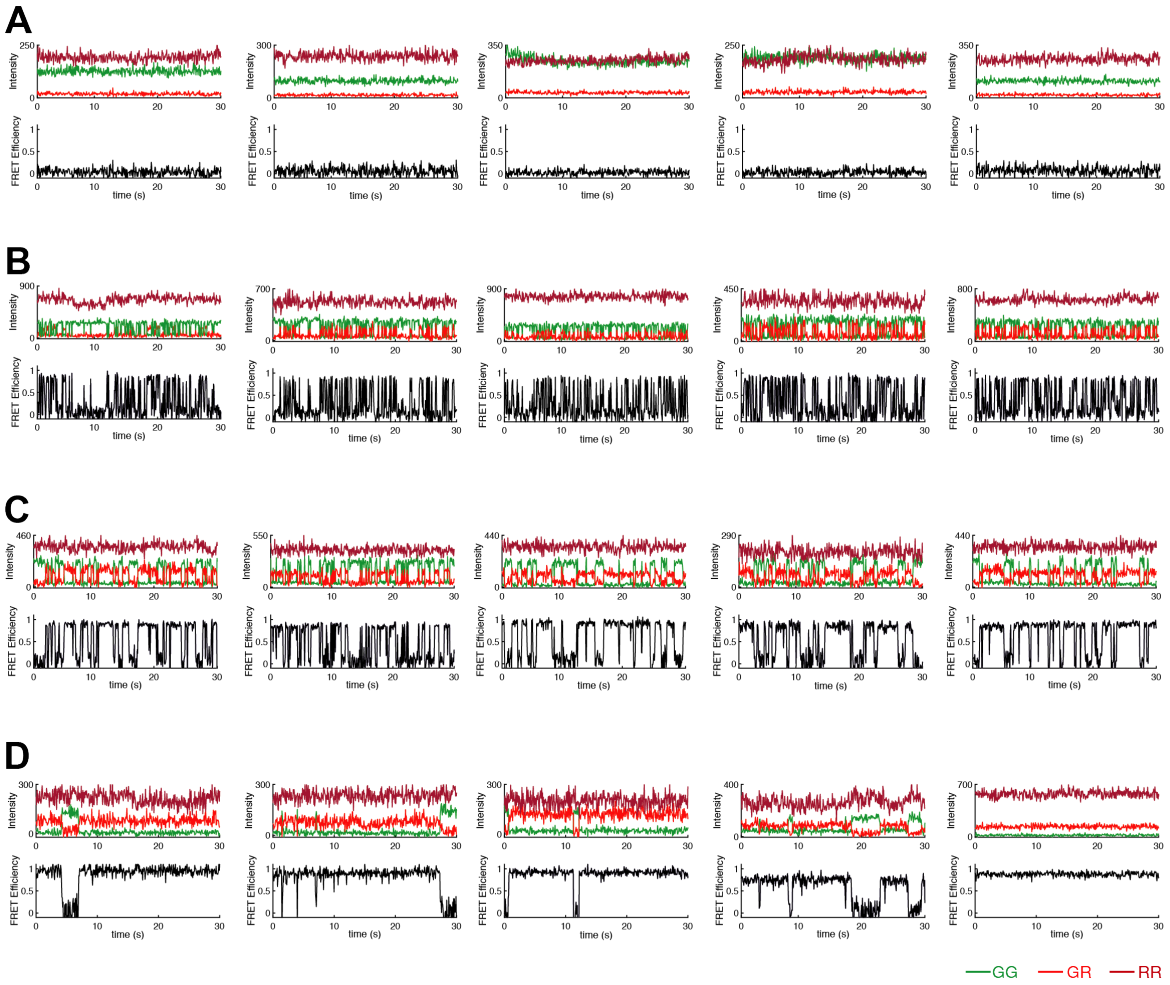


Figure S9. Randomly selected fluorescence intensity and corresponding FRET efficiency traces for surface-immobilized DNA origami platforms with bound DNA double strands without overhangs (blunt ends, **A**), with 4-nt sticky ends (**B**), with 6-nt sticky ends (**C**) and 8-nt sticky ends **(D**) before ligation. The donor signal after donor excitation is shown in green and the acceptor signal after donor excitation (i.e. the FRET signal) is shown in red and the acceptor signal after acceptor excitation (i.e. the ALEX signal) is shown in dark red.

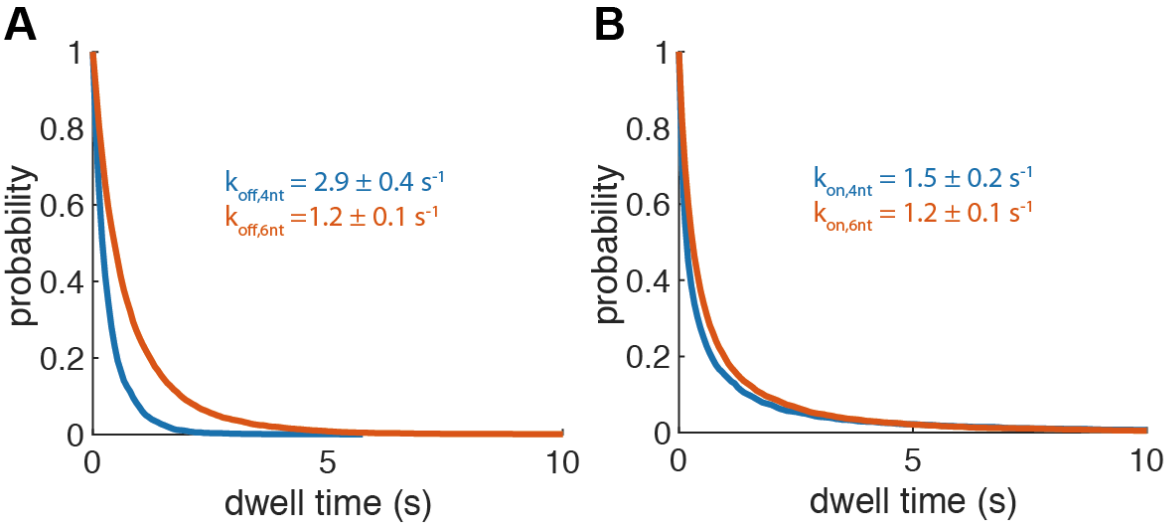


Figure S10. Exemplary dwell time histograms for surface-immobilized DNA origami platforms with 4 (blue) or 6 (orange) nt overhang dsDNAs. **A**) Histogram of dwell times spent in the high FRET state and corresponding off-rates ± standard deviation. The off-rate considerably decreases with increasing overhang length. **B**) Histogram of dwell times spent in the low FRET state and corresponding on-rates ± standard deviation, which are similar for 4- and 6-nt overhangs.

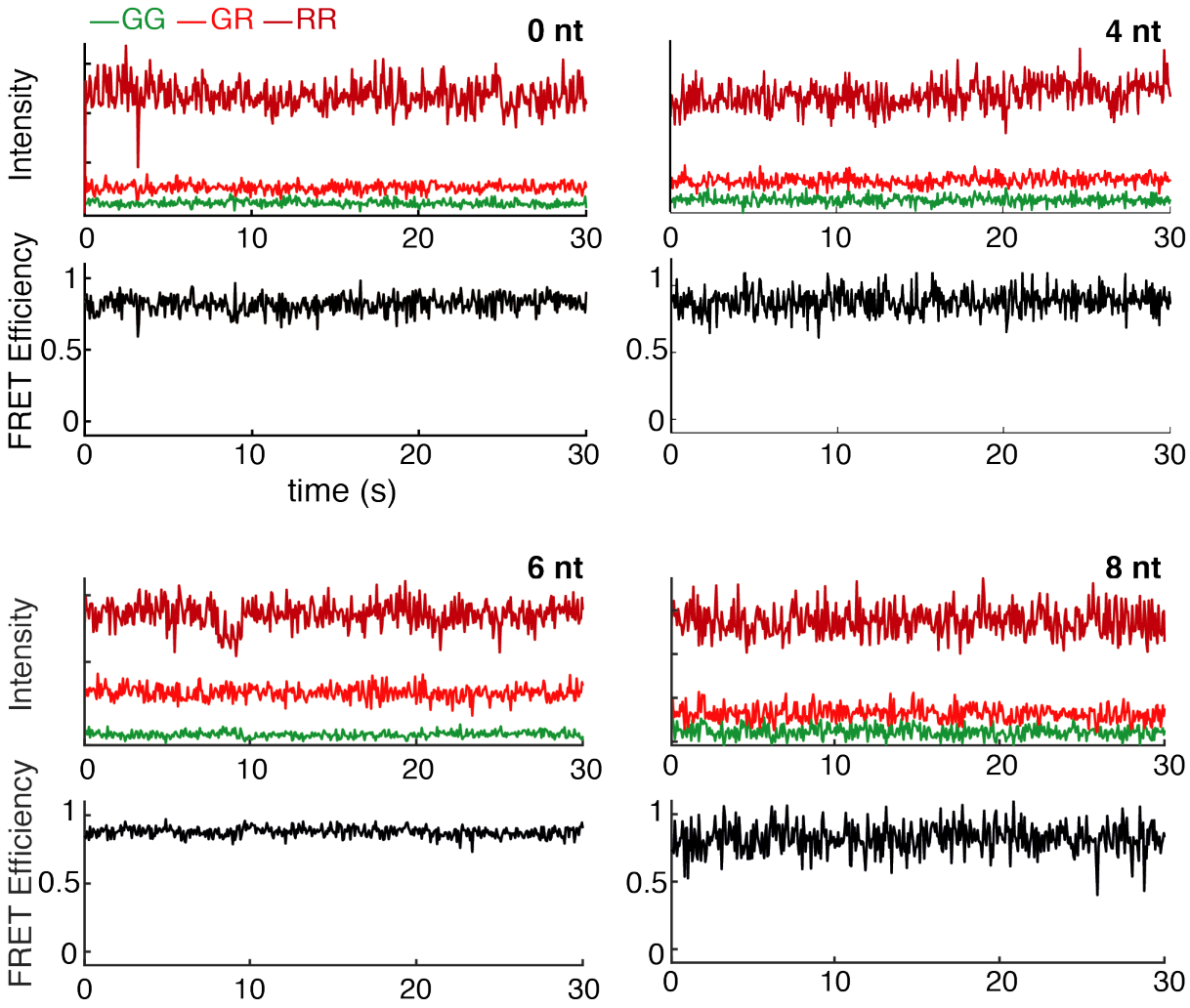


Figure S11. Exemplary fluorescence intensity and corresponding FRET efficiency traces after incubation with the T4 DNA ligase with the donor signal after donor excitation in green, the FRET signal in red and the ALEX signal in dark red. Similar continuous high-FRET traces were found for all overhang lengths, which indicates a successful ligation reaction.

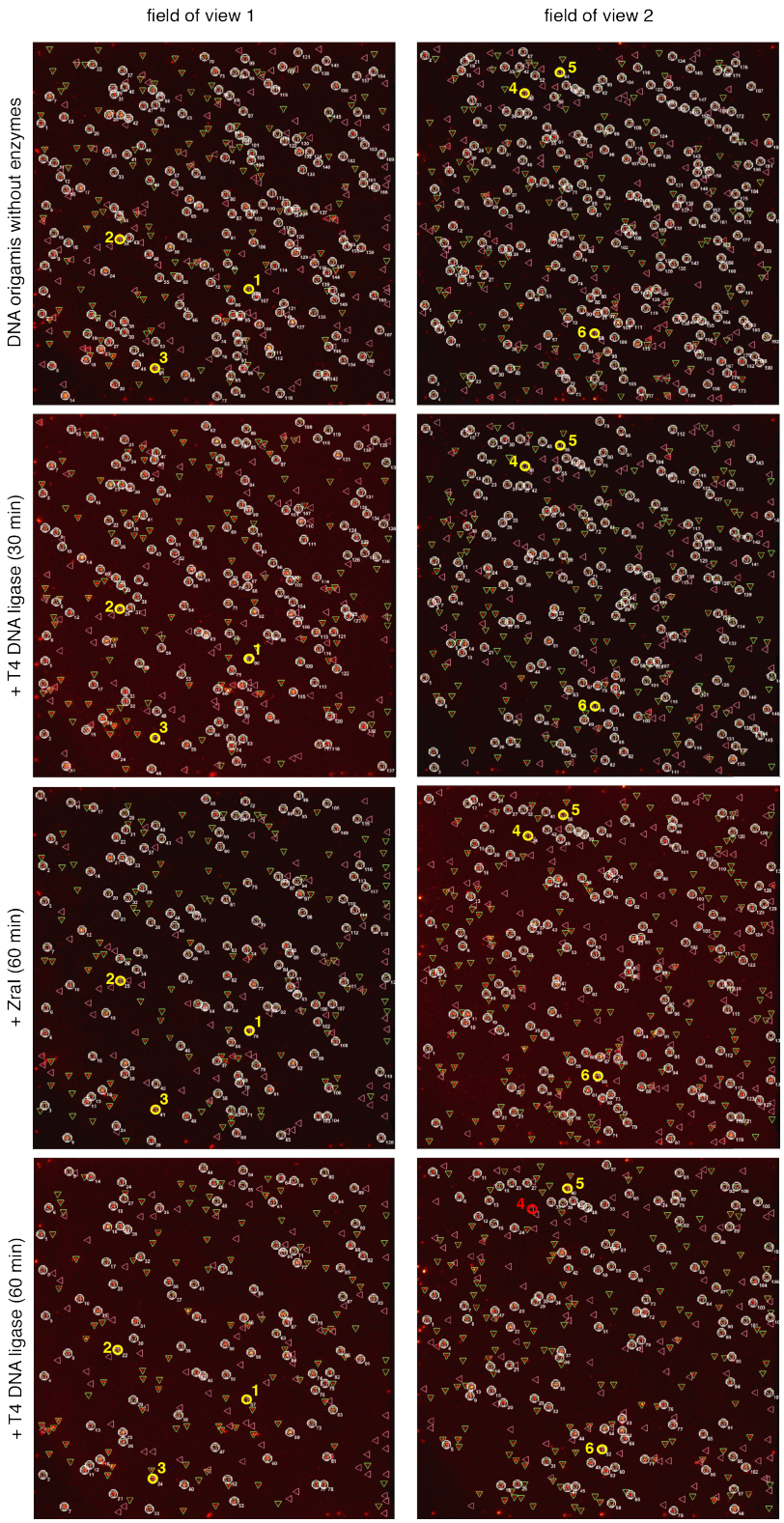


Figure S12. Two fixed field of views for subsequent ligation and cleavage reactions on surface-immobilized DNA origami platforms illuminated by TIRFM. The cumulative images of molecules detected after red excitation (red triangles) or green excitation (green triangles) are shown for each condition. Colocalized molecules are indicated by a white circle and a number. The molecules chosen for the exemplary fluorescence intensity traces in Figure 5 and Figure S12 are marked by a yellow circle and a number from 1-6. For molecule 4, no green signal was detected after the second ligation step. The missing colocalization is indicated by a red circle.

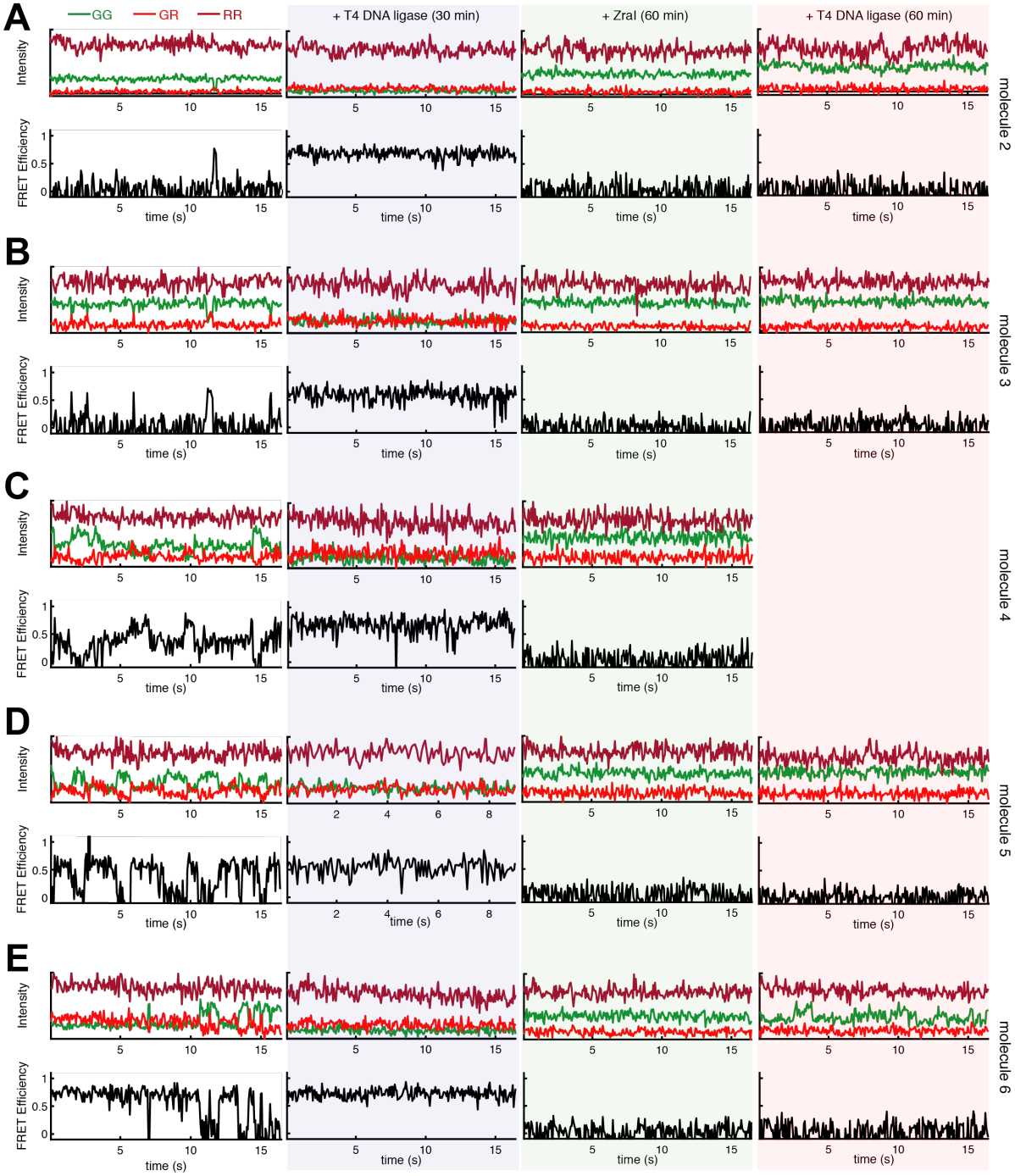


Figure S13. Exemplary fluorescence intensity traces of origami-bound DNA double strands before the addition of the T4 DNA ligase, after incubation with the T4 DNA ligase, after subsequent incubation with the restriction enzyme ZraI and after another incubation step with the T4 DNA ligase. The donor signal after donor excitation is shown in green, the acceptor signal after donor excitation (i.e. the FRET signal) is shown in red and the acceptor signal after acceptor excitation (i.e. the ALEX signal) is shown in dark red. Molecule numbers correspond to Figure S11. In contrast to the exemplary FRET trace in Figure 5C (molecule 1), the molecules presented here show no FRET after the second ligation step.

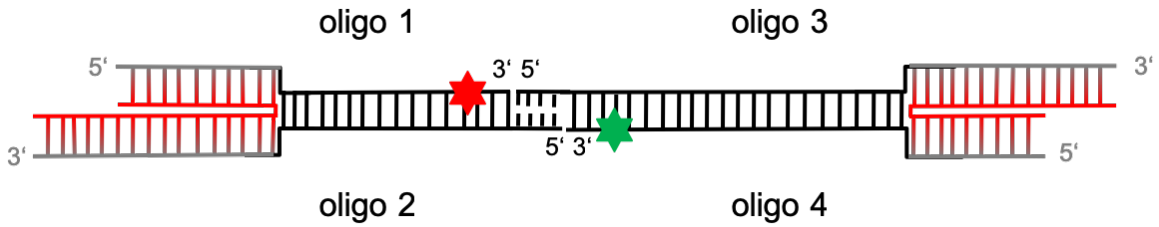


Figure S14. Schematic presentation of the fluorescently-labeled double-stranded DNAs. Oligos 1 and 2 form one DNA double strand, which is labeled with Atto647N (depicted as a red star) and oligos 3 and 4 form the other DNA double strand, which is labeled with Cy3b (depicted as a green star). Parts of the oligos, which are complementary to the DNA origami scaffold strand (shown in red) are colored grey. A complete list of all oligo sequences for overhangs of 0, 4, 6 and 8 nt is given in table S1.

Movie S1. A movie of the first 2 x 10^7 time steps of the coarse-grained MD simulation, corresponding to a simulation time of 303 ns and an approximate physical timescale of 100 μs (see Materials and Methods of the main text). To highlight the global motions of the origami, the trajectory was smoothed using a window of 10 time steps and rendered every 5 time steps.

Movie S2. Excerpt of Movie S1 rendered at every time step and without smoothing to highlight the local structural fluctuations within the origami.

Table S1. The sequences used for the fluorescently-labeled DNA double strands. Oligos 1 and 2 form one DNA double strand and oligos 3 and 4 form the other DNA double strand (Figure S14). Parts of the sequences, which are complementary to the DNA origami scaffold strand, are shown in grey, while complementary overhang sequences are represented by bold letters.

name	Sequence 5'→3'
oligo 1, 0 nt	TTCTGACCTGAAAGCGTAAGCTATGTATGCTCCCATGCTAAATACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATGCCGTG (79nt)
oligo 2, 0 nt	PhosphateCACGGCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAGCATACATAGAATACGTGGCACAGACAATATTTTTGAAT (88 nt)
oligo 3, 0 nt	PhosphateAGATAATTCGTCGGTATGCTATGTATGCTCCCATGCTAAATACAACAAGCGGCGCGCGTGCGCATCGTAACCGTGCATCTGCCAGTT (87 nt)
oligo 4, 0 nt	GGTCACGTTGGTGTAGATGGACGCGCGCCGCTTGTTGTATTTAGCATGGGAGCATACATAGCATACCGACGAAT(T-Cy3b)ATCT (79 nt)
oligo 1, 4 nt	TTCTGACCTGAAAGCGTAAGCTATGTATGCTCCCATGCTAAATACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATGC (75nt)
oligo 2, 4 nt	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAGCATACATAGAATACGTGGCACAGACAATATTTTTGAAT (88 nt)
oligo 2, 4 nt for control dsDNA	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAGCATACATAGCTTACGCTTTCAGGTCAGAA (79 nt)
oligo 1 (+2nt), 4 nt	TTCTGACCTGAAAGCGTAAGGTCTATGTATGCTCCCATGCTAAATACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATGC (77nt)
oligo 2 (+2nt), 4 nt	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAGCATACATAGACAATACGTGGCACAGACAATATTTTTGAAT (90 nt)
oligo 1 (+10nt), 4 nt	TTCTGACCTGAAAGCGTAAGGCGTATCAGTCTATGTATGCTCCCATGCTAAATACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATGC (85nt)
oligo 2 (+10nt), 4 nt	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAGCATACATAGACTGATACGCAATACGTGGCACAGACAATATTTTTGAAT (98 nt)
oligo 1 (-2nt), 4 nt	TTCTGACCTGAAAGCGTAAGATGTATGCTCCCATGCTAAATACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATGC (73nt)
oligo 2 (-2nt), 4 nt	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAGCATACATAATACGTGGCACAGACAATATTTTTGAAT (86 nt)
oligo 1 (-10nt), 4 nt	TTCTGACCTGAAAGCGTAAGTCCCATGCTAAATACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATGC (65nt)
oligo 2 (-10nt), 4 nt	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTATTTAGCATGGGAAATACGTGGCACAGACAATATTTTTGAAT (78 nt)

oligo 1 (hairpin), 4 nt	TTCTGACCTGAAAGCGTAAGCTATGTATGCGCATTCGGGTGA CGAATGCCCATGCTAAATACAACAAGCGGGCGTGAGACAGA GGAT(T-Atto647N)ATGC (92 nt)
oligo 2 (hairpin), 4 nt	Phosphate**CACG**GCATAATCCTCTGTCTCACGCCCGCTTGTTGTA TTTAGCATGGCGTCGAGACTGCCTCGACGGCATACATAGAAT ACGTGGCACAGACAATATTTTTGAAT (105 nt)
oligo 3, 4 nt	Phosphate**CGTG**AGATAATTCGTCGGTATGCTATGTATGCTCCC ATGCTAAATACAACAAGCGGCGCGCGTGCGCATCGTAACCGT GCATCTGCCAGTT (91 nt)
oligo 3, 4 nt for control dsDNA	Phosphate**CGTG**AGATAATTCGTCGGTATGCTATGTATGCTCCC ATGCTAAATACAACAAGCGGCGCGCGTCCATCTACACCAACG TGACC (83 nt)
oligo 4, 4 nt	GGTCACGTTGGTGTAGATGGACGCGCGCCGCTTGTTGTATTTA GCATGGGAGCATACATAGCATACCGACGAAT(T-Cy3b)ATCT (79 nt)
oligo 1 (RE), 4 nt	TTCTGACCTGAAAGCGTAAGCTATGTATGCTCCCATGCTAAAT ACAACAAGCGGGCGTGAGACAGAGGATAATGG**ACGT** (79nt)
oligo 2 (RE), 4 nt	PhosphateCCAT(T-Atto647N)ATCCTCTGTCTCACGCCCGCTTGTTGTATTTA GCATGGGAGCATACATAGAATACGTGGCACAGACAATATTTT TGAAT(84 nt)
oligo 3 (RE), 4 nt	PhosphateCGAT(T-Cy3b)ATTCGTCGGTATGCTATGTATGCTCCCATGCTAAA TACAACAAGCGGCGCGCGTGCGCATCGTAACCGTGCATCTGC CAGTT(87 nt)
oligo 4 (RE), 4 nt	GGTCACGTTGGTGTAGATGGACGCGCGCCGCTTGTTGTATTTA GCATGGGAGCATACATAGCATACCGACGAATAATCG**ACGT** (79 nt)
oligo 1, 6 nt	TTCTGACCTGAAAGCGTAAGCTATGTATGCTCCCATGCTAAAT ACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)ATG (74nt)
oligo 2, 6 nt	Phosphate**TCACGG**CATAATCCTCTGTCTCACGCCCGCTTGTTGT ATTTAGCATGGGAGCATACATAGAATACGTGGCACAGACAAT ATTTTTGAAT (89 nt)
oligo 3, 6 nt	Phosphate**CCGTGA**GATAATTCGTCGGTATGCTATGTATGCTCC CATGCTAAATACAACAAGCGGCGCGCGTGCGCATCGTAACCG TGCATCTGCCAGTT (92 nt)
oligo 4, 6 nt	GGTCACGTTGGTGTAGATGGACGCGCGCCGCTTGTTGTATTTA GCATGGGAGCATACATAGCATACCGACGAAT(T-Cy3b)ATC (78 nt)
oligo 1, 8 nt	TTCTGACCTGAAAGCGTAAGCTATGTATGCTCCCATGCTAAAT ACAACAAGCGGGCGTGAGACAGAGGAT(T-Atto647N)AT (73nt)
oligo 2, 8 nt	Phosphate**CTCACGGC**ATAATCCTCTGTCTCACGCCCGCTTGTT GTATTTAGCATGGGAGCATACATAGAATACGTGGCACAGACA ATATTTTTGAAT (90 nt)
oligo 3, 8 nt	Phosphate**GCCGTGAG**ATAATTCGTCGGTATGCTATGTATGCTC CCATGCTAAATACAACAAGCGGCGCGCGTGCGCATCGTAACC GTGCATCTGCCAGTT

	(93 nt)
oligo 4, 8 nt	GGTCACGTTGGTGTAGATGGACGCGCGCCGCTTGTTGTATTTAGCATGGGAGCATACATAGCATACCGACGAAT(T-Cy3b)AT (77 nt)

Table S2. Quantification of the apparent ligation efficiency for DNA double strands bound to the DNA origami platform. Statistics regarding the high-FRET efficiency population used to calculate the fraction of high-FRET events in Figure 2D.

	number of high-FRET burst with E > 0.6	number of total bursts	fraction of high-FRET bursts
0-nt overhang (o/n, 16° C)	449	2342	0.1917
4-nt overhang (o/n, 16° C)	3363	5841	0.5758
4-nt overhang (1 h, RT)	4322	8541	0.5060
6-nt overhang (o/n, 16° C)	2227	5307	0.4196
8-nt overhang (o/n, 16° C)	4171	7370	0.5659
4-nt overhang, -2nt (1 h, RT)	1863	5227	0.3564
4-nt overhang, -10nt (1 h, RT)	1396	4109	0.3397
4-nt overhang, +2nt (1 h, RT)	3007	6850	0.4390
4-nt overhang, +10nt (1 h, RT)	425	2415	0.1760
4-nt overhang, hairpin (1 h, RT)	2010	3905	0.5147

Table S3. Quantification of the apparent ligation efficiency for freely diffusing DNA double strands. Statistics regarding the high-FRET efficiency population used to calculate the fraction of high-FRET events for the control ligation experiments with isolated 4-nt overhang dsDNAs. The apparent ligation efficiency, given by the fraction of high-FRET bursts normalized to the sum of acceptor-only and high-FRET bursts, is directly comparable to the high-FRET fraction of 0.4765 for DNA origamis with bound 4-nt overhang reported in Figure S6, as the same T4 DNA ligase batch was used for the experiments.

	control dsDNA standard concentration (600 pM)	control dsDNA high concentration (20 nM)
number of donor-only bursts ($S > 0.8$)	24433	19026
number of acceptor-only bursts ($S < 0.2$)	17063	12629
number of high-FRET burst ($E > 0.6, 0.2 < S < 0.8$)	7064	10434
number of total bursts	48560	42089
fraction of donor-only bursts	0.5031	0.4520
fraction of acceptor-only bursts	0.3514	0.3001
fraction of high-FRET bursts	0.1455	0.2479
normalized fraction of high-FRET bursts	0.2928	0.4524

Table S4. Quantification of the apparent ligation efficiency for DNA double strands with restriction enzyme recognition site bound to the DNA origami platform. Statistics regarding the high-FRET efficiency population used to calculate the fraction of high-FRET events for Figure 5B.

	number of high-FRET burst with E > 0.5	number of total bursts	fraction of high-FRET bursts
before ligation	1564	7191	0.2175
after ligation	3040	7744	0.3926
after ligation + ZraI	1229	3986	0.3083
re-ligation after ZraI	1244	3220	0.3863

Table S5. List of all staple sequences used to assemble the DNA origami two-layer frame with opening based on the p7560 scaffold. Endcaps (staples at the ends of the DNA helices) are elongated with T_4-spacers at their 3'- and 5'-termini to reduce aggregation of the folded DNA origami structures. Two endcaps are modified with biotin for surface-immobilization. Start and end positions are denoted as "helix number [basepair number]". The DNA origami was designed with the program caDNAno[3].

name	Sequence 5`→3´	start	end
Oligo 1	ATTGACCACCAGGAGGAAGGTTATCTAAAATATCTTGTAA	23[84]	24[100]
Oligo 2	GTGAGAAAGCAAAGAACGGAACAACATTATTACAGGTAGATCA	16[231]	23[211]
Oligo 3	ACATGATTAAGACGCTGAGAAGAGTCAATCCA	19[100]	21[83]
Oligo 4	GGCATCCGCTCACAATTCCCTTGAATCCG	1[236]	36[228]
Oligo 5	CAATATGAAGCATTAAATCAGTTGACCAGACG	16[199]	21[199]
Oligo 6	TAAAAAGAGAATCGATGAACATTGTATAGAAT	16[275]	25[283]
Oligo 7	AAATCAGTTCAGAAAACGAGAATGACCATAGACTGGATTT	19[292]	18[276]
Oligo 8	GCTGAAATCAGGTCTTTACCCTGACTATAGTT	19[260]	21[243]
Oligo 9	ATTGACAGGAGGTTGAGGCACAAA	39[284]	38[300]
Oligo 10	CCTAAATAAAGATAACGGATTCGCCTGAACCT	22[139]	19[131]
Oligo 11	GCTAAGCGAACACATCGGGCGAGGCATGATACATAACGCCAAACTGA	19[149]	22[140]
Oligo 12	AATCGATTATTCGCTCATGGAAATACAGAA	31[52]	6[60]
Oligo 13	CAACAAGAACCGAACTGACCAACTTTGAAAGAGGATATTAGACCAATTCGA	16[187]	16[168]
Oligo 14	ACAAAGCCCCCTTATTAGCGTTTGCCATGTGAATTAACCC	0[123]	1[139]
Oligo 15	AAGGTCAACAGTTTCAGCGG	33[284]	34[272]
Oligo 16	TGTAGCAAATCGGCCTCTTT	4[71]	34[68]
Oligo 17	TCAACGTCAGATAGTAACAGTACCTTTTCAGA	22[172]	19[163]
Oligo 18	TACATAAAACTATAAAAATTTTTAGAACAAAC	24[299]	17[291]
Oligo 19	TATAATGATGAATATAATCCTGATTGATCA	21[52]	16[60]
Oligo 20	AGGAACAGCCCTCATAGTTAGCGTAACGGTTGTGAAGACCTCCT	36[259]	3[279]
Oligo 21	AAAATGTTTAAATCAAAATATAATGCTG	21[264]	19[275]
Oligo 22	TAGGGCCCTGACCCGCGACCTGCTCCATGTTACTTAAAAGGCTCAAAAGG	16[251]	16[232]
Oligo 23	GAGTATAAGTATTTTTGTCGTCTTTCCATTTT	3[292]	34[300]
Oligo 24	ATCGACGGAGATTGAGATGGTTTAATTTGAT	15[292]	22[301]
Oligo 25	CCACCACCAATCCCTTATAAATCAAAAGAATAGCC	39[176]	0[156]
Oligo 26	AGAAATAATAAGAGCAAGAAACAATGAACGCCAGAAAAGGGATT	36[131]	3[151]
Oligo 27	CCGCACCGCCTGGCCCTGAGAGAGTTGCGCGGTTTGTCTGAATTAGAGGCTG	39[276]	37[263]
Oligo 28	GCAAGGAACCGCCTCCCTCAGAGCCGCCTACTGGTATCGG	0[187]	1[203]

Oligo 29	CCAGAACAACTTAACAACTAAAGGAATTTCCA	5[292]	32[300]
Oligo 30	CCATCTGTAAAGCCTGCAGTGCCCTAACGGGGTCAGTGCCGGT	3[184]	38[173]
Oligo 31	TTGCGCTAACGAGCGTCTTTCGGGAGGTATTC	32[67]	9[75]
Oligo 32	GTTTAATTTAGGATAAAGTACCGACAAAAGGTAAATAGGAGCAATTATCAT	16[123]	16[104]
Oligo 33	AGAGGAGGAAGCCCGAAAGACTTCAAATATCA	19[196]	21[179]
Oligo 34	TGACAATGGTCACCAGCCTCAGAACCGCCACCCCCC	4[223]	37[211]
Oligo 35	CAGAGTCCACGCTGGTTTGCCCCAGCACATTAATGTGGCTTTGAACCTAT	39[244]	37[232]
Oligo 36	TCGGTTCCTGTATTTTTTAACCAATAGGTTGT	28[299]	13[291]
Oligo 37	GAAGACTATATGTTACTAGA	18[59]	22[56]
Oligo 38	TAGGCTTTGCCCATAATACATTTGAGGATTTAGAAGCAGA	23[148]	24[164]
Oligo 39	TAACCTAATACGAATCTAATAGTAGTTATT	21[180]	16[188]
Oligo 40	TAAATAGCAACGGCTACAGACTGAGGCTCGGCGGATCGACGACA	27[284]	9[279]
Oligo 41	AATGGAAAAGAGGCGACAAGACAAGCGTTAAATAAGAATACTTA	19[88]	23[83]
Oligo 42	AATAGAACAATAATCGTTAGAACACCGAGTAAAAGAGT	37[104]	4[96]
Oligo 43	TCTTCCTGAACAATTACGCAGTATGTTAGCA	4[59]	37[50]
Oligo 44	ATGAAAATTGTTCAACGCGCGGGGAGAGAGCA	37[244]	0[252]
Oligo 45	GAGTGTTGTTCCAGTTAAATCGGATCACCGTCCAGCGCCA	0[143]	37[135]
Oligo 46	TTGGATAAAGCTTGACCATTAGATACATAATT	22[267]	19[259]
Oligo 47	CGAGAATCAAAATCACCGGAACCAGAG	0[155]	39[175]
Oligo 48	ACGATAAAAAGATTAATACCTTTAATTGCTCC	21[200]	19[215]
Oligo 49	TATAATCATAAGCCCAGGAAACCGAGGAAACGAATT	4[127]	37[115]
Oligo 50	GGGAGCTAAGAAAGGATGGTTTACACCGACTTGAGCCATTTCGG	3[120]	39[115]
Oligo 51	ATAACCGATGAGGGGATGACCGTAATGGGATA	30[271]	10[256]
Oligo 52	TTTTAACGCCAGGGTTCTAAGTGATCT	5[273]	35[283]
Oligo 53	TGGCGTCTTTAATGCGCGAAAGATAAAATTTC	9[76]	28[68]
Oligo 54	GCTTTAAAGATTATCAGGTCATTGCCTGAGAGTCTGGTCG	23[244]	24[260]
Oligo 55	ATCGATTATTCAGAATAATGGAAGGGTTACAA	21[84]	16[92]
Oligo 56	TATAACGTGCGCTAGGTTATTTTGGCAAAATCACCAGTAGGTCA	3[88]	39[83]
Oligo 57	CATAAGGGAAGGTGCCGTAAAGCACTTGGA	37[116]	0[124]
Oligo 58	GAATGAGATAACCCACAAGAATTGAGTGTGAGGCTCAGAGC	36[99]	3[119]
Oligo 59	ATTATGACCCTGATTCCCAATTCTGCGATTTT	17[292]	19[291]
Oligo 60	TGAACCTTCATCAAGAGTACAT	24[163]	22[173]
Oligo 61	CTACGTTAATATCCCCAAAAACAGGAAGGGTA	26[299]	15[291]
Oligo 62	TAAACCAGTGAGGATCCCCGGGTACCGATAAT	38[299]	3[291]
Oligo 63	TACCAACACCGCGAACCACCAGCAGACTGA	27[52]	10[60]
Oligo 64	GCAAAAGATATAGTCATTTTTGCGGATGGCT	21[232]	19[247]

Oligo 65	TTTTAGGTGAGGCGGAAAAATCTAACC	12[78]	26[68]
Oligo 66	CGACGATCGCACCGTCGGATTCTCCGTGGTCT	30[299]	11[291]
Oligo 67	CGAGAAAGCGAAAGGAGCGGGCTTTCCTAACG	1[108]	36[100]
Oligo 68	AAAAGCGATCGGAGGCAAAGCGCCATTCCTCA	32[299]	9[291]
Oligo 69	TAAATTGTGAGCAAACTGCAATGC	24[271]	16[264]
Oligo 70	TATTCTGCCTCAGAACGGCTGAGAATTGTCAACC TTA	37[233]	4[224]
Oligo 71	GGCCAACGAGGGGACTTTTTCATGAGGAGCCA	11[292]	26[300]
Oligo 72	AGTTAATGCTCAGAGCCGTCGGTGAAATAACCC CGCTTCT	37[200]	4[192]
Oligo 73	ACATAAATCCTGTGTGAAGTATTATACCGTTCCA GTAAGCCCCT	3[248]	39[243]
Oligo 74	TTTTAAATAATTCGCGGAACAAATGCA	11[273]	29[283]
Oligo 75	TTTTACGGGCAACAGCTGATTCTT	0[307]	1[291]
Oligo 76	AGTGAATATTGCTTTGATATATTTGGTTTGAAAT ACCGACCAAC	19[120]	23[115]
Oligo 77	GGCTTTGTATCATCGCCTGAAGGCAAAAAGCTTT T	23[284]	14[273]
Oligo 78	GGAATTGAGTAAGGGTGCCTGCGCTCACTGCCC GATCG	37[171]	0[188]
Oligo 79	CTGAGTAACTCAGAGCGAAGAAAAATCTACGTT AATAAAATAAG	16[263]	23[243]
Oligo 80	GGAATGCTCATTCAGTGAACGAA	24[227]	22[236]
Oligo 81	AATGTTACCAACCAGTTACAAAATAAACAAAC	8[59]	33[51]
Oligo 82	GGTCAGGACATCAATTCCTCGTTTAGATTTAGGA ATACCAATCT	19[184]	23[179]
Oligo 83	GATGCAAATAGTGAATTTGAATTACCTTTTTT	21[72]	19[87]
Oligo 84	GAAACGGAAGCATAAAGTGTAAGCAACTCACC	1[204]	36[196]
Oligo 85	AGAAAACTAGCTTAAAATCAATATATGTG	21[104]	19[119]
Oligo 86	TTTTGCGATGGCCCACCCAT	1[55]	38[68]
Oligo 87	ATCCCCTAAAAACACGACCAGTAATACTGA	29[52]	8[60]
Oligo 88	TAGCGGTATTCTCTTAAATCAAGATTAGCCTG	10[59]	31[51]
Oligo 89	GATAGCCGAGGGCGACCGGGGAAAGGGAGCCC	36[151]	1[151]
Oligo 90	ACACTGATTAGCGCTACAGGGCGCGTCTGCGCG T	35[52]	2[56]
Oligo 91	ACCCCCATGTACCGTAACACTGAGTTTCTCCCGC CAGGCACGAA	36[195]	3[215]
Oligo 92	AATCTATTATAGGAACTCATTTTCAGGG	4[191]	36[180]
Oligo 93	ATCGCGAGAACAAGCAAGCCCATGTAGAAAATT TT	28[67]	13[78]
Oligo 94	ATCATAGGTCTGAGAGGTTA	20[63]	21[51]
Oligo 95	TGACCGTCAAATTTATCTAGCTGATAAATTAATG GGGA	23[180]	24[196]
Oligo 96	CCTTAAGTCAGAGGGTAATTATAGCAGCTGCTTT T	36[67]	5[78]
Oligo 97	CAATTTTTGTTTAGAGAATAACATAAAACTGA	6[59]	35[51]
Oligo 98	GAGCGCTACACGCAAACGAGCACG	35[80]	3[87]
Oligo 99	GGTTGGTGGCTCGAATGGATTAGGAAAGCCAGA ATGGAAAGCCG	3[280]	39[275]

Oligo 100	GGGACAACCATCGCCCACGCTTGTATCGGGCTTTT	29[284]	8[273]
Oligo 101	TATTGCACTCATTAGGAATCATTACCGCGCTT	12[59]	29[51]
Oligo 102	TTTTTGGTAATATCCCTACATTTTAAT	6[78]	32[68]
Oligo 103	TTGGATGGTCATAGCTGTTTCATTTCTCACTC	1[268]	36[260]
Oligo 104	TAGCTCAACATGACGAGTAGATAG	19[276]	21[283]
Oligo 105	CTGTCCATATATCAGAACCCAAAAGAACTGGCGAAT	4[95]	37[83]
Oligo 106	CGTCTCATTATACCAGTCAGAGTAGTAAATTT	21[284]	16[276]
Oligo 107	CAACGTAGCGCGTTTTCATCGGCATTTTGGGAATGGGT	0[91]	1[107]
Oligo 108	ACACCAACCTAAAACGAAAGGGCTTTGACAATTTT	25[284]	12[273]
Oligo 109	AAGACAAAAACAAAGTGCCGATTATCCTGAGAAGTGTTTT	37[136]	4[128]
Oligo 110	AAAAAACCGCCACCCTCAGAGCCACCAGTCATACAAATC	0[219]	1[235]
Oligo 111	GTCAGAAGCAAAGCGGATTGCATCAAAAACC	19[228]	21[211]
Oligo 112	TATAGGGGCACACAACCTATTTCGTGATGATACAGGAGTGACCC	3[216]	39[211]
Oligo 113	CCACTACAAACTACAACGCCTGTAGCATTAGTGCTCGCATTTC	36[227]	3[247]
Oligo 114	TTCATCCTCATTATTAGCGG	1[292]	37[295]
Oligo 115	AAAGAGCCCGGAATAGGTGTCAAGAGAATCGT	35[284]	2[276]
Oligo 116	CAACTCGTAGGTTTAACTAATGCAAGTAAGAG	16[167]	21[167]
Oligo 117	TCTGAACGGTAATAGCAATAGTAAGCA	4[162]	36[152]
Oligo 118	AATCGCGCCAGGGTGGTTTTTGCCCTTCCAGC	2[275]	39[283]
Oligo 119	CCGATTTACTCACATTATTGACGGGATTGAGGGAG	1[152]	37[170]
Oligo 120	AAAATTGGGGCGAAATCATACAGGCAAGGGCC	21[212]	16[220]
Oligo 121	AGAAAGAATCGCGTCCAGACGACGACAA	16[91]	24[80]
Oligo 122	ATTCCATCCAATCGAGCTGAAAAGGTGGTTAG	22[203]	19[195]
Oligo 123	GACTGTCAAAGGGCGAAAAACCGTCTATCAGGTTTT	39[84]	0[55]
Oligo 124	TGAAAAGTGTAGCGGTCACGACTATGGTGACT	1[76]	36[68]
Oligo 125	TTTTAAATATTTAAAAACGCCATGGAC	13[273]	27[283]
Oligo 126	TTAGACAGATGATACCTTAAGAAAAGCTATCT	3[152]	35[167]
Oligo 127	CAACACTATCGCGTAAGCAAACTCCAACA	21[168]	19[183]
Oligo 128	AAGCAAGAACGCGAGGCGTTGTTTTTATCAGTTTT	30[67]	11[78]
Oligo 129	TAGCACGCAAAGACACCACGATGATTAATGCTTTGATTAACCGT	38[67]	4[72]
Oligo 130	AATTAAGGAGCGGAATTATCTTTGGATTAATCATAATAAATGCT	15[76]	21[71]
Oligo 131	ACCTTAATATCCAATAATCGGCTGTCTTCAAG	14[59]	27[51]
Oligo 132	ATTAAGAAACAAAATTGCGTAGATTTTCATTA	21[148]	16[156]
Oligo 133	TATTTAAAGCCACTAATGCAGAACGCGCAAGA	16[59]	25[51]
Oligo 134	AATCCATCCTAATTTACGAGTAAACAACAAGG	26[67]	15[75]
Oligo 135	TTCTTTTCAATTACCTGAGCAATTTCATTTATCAAA	17[76]	20[64]

Oligo 136	AAGTGCGCTGGCCCATCACCCAAATCAAACTC	37[84]	0[92]
Oligo 137	TTCTCATATTTAACAACGCCGT	24[99]	22[108]
Oligo 138	TTTTGATAATTTTCATTAGCGAGAGGCTTTT	19[216]	21[231]
Oligo 139	AAGTGTCCTTCCACAGGGTTTAGTACCGCCACAAAC	4[255]	37[243]
Oligo 140	ATAGGCCCTTTGACAGTGCGCCCTGGAGTGAC	36[179]	4[163]
Oligo 141	AAACGAAGGCACTAAAACACTCATCTTTAAAG	13[292]	24[300]
Oligo 142	AAATAATTGCGTTAATGAGTGAGCTAAGAGCTTGAATTCAACCAAATTATT	38[172]	38[152]
Oligo 143	CTAATTAGCAAAACCTGTTTAGCTATAGAG	22[235]	19[227]
Oligo 144	TCAGTCCTGTTTGATGGTGGTTCCGAACTTTCCAGATAAGTTTGTATAAAC	39[212]	37[199]
Oligo 145	AATCCTGGCTGACGGTGTACAGACCAGGCCAGTAATGTCAATAGGAACGTTA	16[155]	16[136]
Oligo 146	TAAAGCCGGCGAACGTGGCGAACAGGAGTACC	1[140]	36[132]
Oligo 147	TTTGCGGAAGAACCTATAAATAAGAGAACGCG	16[103]	21[103]
Oligo 148	TTGCTGGTCAATAATTAAGCAATAAAGCTGTG	21[244]	16[252]
Oligo 149	TTTTGCATCACCTTGACAGTTGAATGT	14[78]	24[68]
Oligo 150	AGACTCCTATCACCGTCGAACTCTTTCATGCGCACGACTT	37[264]	4[256]
Oligo 151	TCATGAGTCCACTATTAAAGAACGTGGGTTTTTTGTAGAGCCATCACAATC	39[116]	37[103]
Oligo 152	AAAACAAATATAGCCAGCAGCAAATGTCAG	25[52]	12[60]
Oligo 153	ACCGGATATTCATTACCCAAAAAG	24[195]	22[204]
Oligo 154	TCAGACGCTCAACAGTAGGGAACACCGGATAC	24[67]	17[75]
Oligo 155	TTAATTTTAAAACAGAAATTTAATTAGTTAATTTCATCTTAGGA	16[135]	21[147]
Oligo 156	GGAGAACAAAGCCGAGGCGCAGACGGTCAATCATAACCGGAGAGATCACCAT	16[219]	16[200]
Oligo 157	GTATCGGCGCCATTCAGTTT	9[280]	31[283]
Oligo 158	TTAGCGAAGGCTATTACAACAGAGATAGAACCC	29[80]	9[96]
Oligo 159	GGAAAATGACAAGTTAAAGGCCGCTTTTAGCA	9[292]	28[300]
Oligo 160	ACAGAACGTCAAAAATGAAACCAGAGCCTGATTTT	34[67]	7[78]
Oligo 161	TACCGAACAAGCCCATACGCTCGGCCCTG	35[168]	3[183]
Oligo 162	TTTTTGCGCAACTGTTTAAGTTGTAGA	7[273]	33[283]
Oligo 163	ACGTACAGTCAAGGTAGCTATTTTTGAGAGATCTACGCC	23[212]	24[228]
Oligo 164	AAGGAAAGGCTCGCTTGCTTTCGAGGTGGCGC	7[292]	30[300]
Oligo 165	TCAAAATACCAAAATTATTTGCACGTAAAA	21[116]	16[124]
Oligo 166	GATTATTACCGAGAAGAACTCAAACTTACT	33[52]	4[60]
Oligo 167	TTAGAGCCAAGAGAATCAGAGGCATTTTCGAGCGCA	15[128]	23[147]
Oligo 168	ATGTGAGTAACCTAACAACTAATAGA	23[116]	15[127]
Oligo 169	AGCGGCCGCCACCAGAACCACCACCAGAGCGCAGTCCGTA	0[251]	1[267]
Oligo 170	TAGAGCTTTTCGCAAACAGAGGGGGTAATAGT	19[248]	21[263]
Oligo 171	ATCACAAAAGGAGCCTTTAAAGTGAGAAGGTTTTT	31[284]	6[273]
Oligo 172	CTGCATACGAGCCCTGTCGTGCCAGCTGGGCG	37[212]	0[220]

Oligo 173	GTGACCATATCAAGTTACAAAATCGCGCCAGT	22[107]	19[99]
Oligo 174	TTTTCGCTCAATCGTAAAGGGACTTTG	8[78]	30[68]
Oligo 175	AGTTAAATCGGTTGTACCAACCTCATATATTG	18[275]	23[283]
Oligo 176	CCGGTTTAATTCGAGCTTCAATTAATTAATTTTC CCTCTGTAAATCGTC	19[164]	19[148]
Oligo 177	GCTATCACGACGATGTGCTGCAAGGCGATGGG	34[299]	7[291]
Oligo 178	CATTAAAGCTTTTCATATAGGGTT	38[151]	0[144]
Oligo 179	AAATGAGAAACACCAGAACGGACG	24[259]	22[268]
Oligo 180	TAGCGACAGAATCAAGTTTGCCTTTAGCCACCAT TATACG	39[52]	1[75]
Oligo 181	TGCTTTAGAATCCTTGAAAACATAGCGATTTT	19[132]	21[115]
Oligo 182	ACATATAAAAGAAAAGGCCGGAAACGTCATCAG	37[51]	39[51]
Oligo 183	AAAAGCCTTACCAGTACCTGATTTCAGTTGGCAA ATCACTGA	22[55]	14[60]
Oligo 184	TTTAAGAACTGGCCAATACTGCGGAATCGGCTT	22[300]	20[300]
Oligo 185	GGTTTTGCGGGTTGATAAACAGGGCCAGGGTGG ATGTTCTTTTC	37[296]	5[291]
Endcap 1	TTTTACATACATAAAGGTGGCAAACGTAGAAAA TTTTT	37[28]	36[28]
Endcap 2	TTTTGCCACGCTGAGCAAACCCTCAATTTT	13[33]	14[33]
Endcap 3	TTTTTCAGCAGCGAAAGACGCGGGATCGTCACC CTTTT	28[323]	29[323]
Endcap 4	TTTTTTCATTCCATATAACAGTTGTAATACTTTTG CGGGAGAAGTTTT	18[328]	17[328]
Endcap 5	TTTTATAGCAGCACCGTAACCAATGAAACCATCG TTTT	39[28]	38[28]
Endcap 6	TTTTAGATAAGTCCTGAACCTGTTTATCAACAAT TTTT	25[28]	24[28]
Endcap 7	TTTTGGTAAAATACGTAATAGTTTCCATTAAACG TTTT	26[323]	27[323]
Endcap 8	TTTTAACAGGAAAAATACATTGGCAGTTTT	7[33]	8[33]
Endcap 9	TTTTCCCAATCCAAATAAGAGCCATATTATTTAT TTTT	33[28]	32[28]
Endcap 11	TTTTATTTTCTGTATGGGAGACGTTAGTAAATGA TTTT	34[323]	35[323]
Endcap 12	TTTTATTCACCAGTCCATCGCCATTATTTT	9[33]	10[33]
Endcap 13	TTTTGCTTAATGCGCTAATAACATCATTTT	3[33]	4[33]
Endcap 14	TTTTAGAAAACAAAATTAATTACATTTAACAAAA	19[33]	18[60]
Endcap 15	TTTTCAGCTACAATTTTATTTGCTATTTTGCACCT TTT	31[28]	30[28]
Endcap 16	TTTTTCAATATCTGGATCAGATGATGTTTT	15[33]	16[33]
Endcap 17	TTTTTTCACGTTGAAAATCGCGAATAATAATTTT TTTT	32[323]	33[323]
Endcap 18	TTTTAATCAGATATAGAAGGCCCAATAGCAAGC ATTTT	29[28]	28[28]
Endcap 19	TTTTGAATCCCCCTCAAATTCATAAATATTCATTT TTT	20[323]	21[323]
Endcap 20	TTTTAGACGGGAGAATTAAACAGGGAAGCGCAT TTTTT	35[28]	34[28]

Endcap 21	AACCACCACACCCGCCGCTTTT	2[55]	2[33]
Endcap 22	TTTTAATGTGAGCGAGTAACAACCTCCAGCCAGC TTTCCGGCACTTTT	10[328]	9[328]
Endcap 23	TTTTTGAATTACCTTATGCCAACTTTAATCATTGT TTT	22[323]	23[323]
Endcap 24	TTTTGAACGGGTATTAAACTCCTTATCATTCCAA TTTT	27[28]	26[28]
Endcap 25	TTTTCGTTATACAAATTCTGTTTAGTATCATATGT TTT	23[28]	22[28]
Endcap 26	TTTTGCAATTCATCAAACAAACATCATTTT	17[33]	18[33]
Endcap 27	TTTTATTTTTGTTAAATCAGCTCAGCCAGCTTTCA TCAACATTATTTT	12[328]	11[328]
Endcap 28	TTTTCGCTTCTGGTGCCGGAAACCTGCGGGCCTC TTCGCTATTATTTT	8[328]	7[328]
Endcap 29	TTTTTACCAAGCGCGAAACGACCCCCAGCGATTA TTTT	24[323]	25[323]
Endcap 30	TTTTCTTGCCTGAGTCCAGCCATTGCTTTT	5[33]	6[33]
Endcap 31	TTTTTTGATACCGATAGTTAATTTCTTAAACAGC TTTT	30[323]	31[323]
Endcap 33	TTTTCCTTTATTTCAACGCAAGGAGCATGTCAAT CATATGTACCTTTT	16[328]	15[328]
Endcap 34	TTTTAAAATACCGAACCTGCAACAGTTTTT	11[33]	12[33]
Endcap 35	TAAAATGCAACTAAAGTACGGTGTCTGGAAGTT TTT	20[299]	19[328]
Endcap 36	TTTTCGCCAGCTGGCGAAAGGGGGTTGTAAAAC GACGGCCAGTGTTTT	6[328]	5[328]
Endcap 37	TTTTTCCGGCTTAGGTTGGACTACCTTTTTAACCT TTT	21[28]	20[28]
Endcap 38	TTTTCCAAGCTTTCTCAGGAGAAGCTTAAGCTAC GTGGTGCTTGTTTT	4[328]	3[328]
Endcap 39	TTTTTTACCTCGATAAAGACGGAG	2[328]	2[304]
Endcap 40	TTTTCCGGTTGATAATCAGAAAAGTTTGTTAAAA TTCGCATTAATTTT	14[328]	13[328]
Endcap 10 with biotin	Biotin-TTTTTTTTTTAAGTGCCGTCGAGATCAGTACCAG GCG GATTTT	36[323]	37[323]
Endcap 32 with biotin	Biotin-TTTTTTTTTCTTGATATTCACAAAGGTCAGACGA TTG GCTTTT	38[323]	39[323]

Table S6. Dye parameters for accessible volume simulations with the AV3-model.

	linker length [Å]	linker width [Å]	R1 [Å]	R2 [Å]	R3 [Å]
dT-Cy3b	20.4	4.5	7.1	5.0	1.5
dT-Atto647N	20.4	4.5	7.2	4.5	1.5

References

(1) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: an R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695–2696.

(2) Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M. Self-Assembly of DNA Into Nanoscale Three-Dimensional Shapes. *Nature* **2009**, *459*, 414–418.

A.2 Paper 2: Structural basis of rotavirus RNA chaperone displacement and RNA annealing

Latest manuscript draft.

Structural basis of rotavirus RNA chaperone displacement and RNA annealing

Jack P. K. Bravo[1,2,†] Kira Bartnik[3], Emma H. Gail[4,5], Chen Davidovich[4,5], Don C Lamb[3], Roman Tuma[1,6], Antonio N. Calabrese[1], Alexander Borodavka*[1,2,3,#]

1. Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK
2. Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK
3. Department of Chemistry, Center for NanoScience (CeNS), Nanosystems Initiative Munich (NIM) and Centre for Integrated Protein Science Munich (CiPSM), Ludwig Maximilian University of Munich, Munich, Germany
4. Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Victoria, Australia
5. EMBL-Australia and the ARC Centre of Excellence in Advanced Molecular Imaging, Clayton, Victoria, Australia
6. Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

* To whom correspondence should be addressed.

† Current address: Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712-1597, USA

Abstract

Rotavirus genomes are distributed between 11 distinct RNA segments, all of which are essential for virus replication. Stoichiometric genome segment selection and assembly is achieved through a series of sequence-specific, intersegment RNA-RNA interactions that are facilitated by the rotavirus RNA chaperone protein NSP2. The C-terminal region (CTR) of NSP2 has been proposed to play a role in rotavirus replication, although its mechanistic contribution to the RNA chaperone activity of NSP2 remained unknown. Here, we use single-molecule fluorescence assays to directly demonstrate that the CTR is required for promoting RNA-RNA interactions and that it limits the RNA unwinding activity of NSP2. Unexpectedly, hydrogen-deuterium exchange-mass spectrometry and UV-crosslinking data indicate that the CTR does not interact with RNA. However, removal of the CTR reduced the RNA release activity of NSP2, suggesting that the CTR is important for chaperone recycling. To further interrogate the role of the CTR, we determined cryo-EM structures of NSP2 and its ribonucleoprotein complexes. These reveal that although the CTR is ampholytic in nature, it harbours a highly conserved acidic patch that is poised towards bound RNA. Together, our data suggest that the CTR reduces the accumulation of kinetically trapped NSP2-RNA complexes by accelerating RNA dissociation via charge repulsion, thus promoting efficient intermolecular RNA-RNA interactions during segment assembly.

Introduction

The ability of RNA to perform diverse roles, including transcriptional and translational regulation, is attributed to its intrinsic ability to adopt an ensemble of alternative conformations due to extensive base-pairing (*1–4*). Similarly, RNA-RNA interactions are favored in cells, yet their formation is governed by a myriad of RNA-binding proteins that allow conformational selection of thermodynamically stable RNA duplexes. Cells deploy a wide range of RNA-binding chaperone proteins whose role is to limit non-specific RNA-RNA interactions that interfere with their functions, while promoting the formation of biologically important RNA contacts. This is particularly pertinent to viral RNAs that can be translated, replicated or packaged (*5–7*). Remarkably, several polyphyletic groups of viruses have expanded their genomes via RNA segmentation, and thus face the additional challenge of selectig and assembling a complete set of distinct RNAs upon viral replication. Amongst these, Rotaviruses (RVs) have eleven distinct genomic segments, all of which are essential for viral replication (*8–10*), epitomizing a particular mechanistic conundrum of a robust RNA selection guided by inter-molecular RNA interactions. To achieve such remarkable stoichiometric assembly of non-identical transcripts, the rotavirus genome encodes the RNA chaperone protein NSP2, which enables efficient RNA selection by mediating a series of intersegment RNA-RNA interactions (*10–12*).

Aside from NTPase and RTPase activities, the RNA chaperone activities of NSP2 include multivalent ssRNA-binding, RNA unwinding, and duplex formation (*13*, *14*). Together, these activities allow NSP2 to facilitate sequence-specific intermolecular interactions between regions of RNA that may otherwise be sequestered by stable RNA structures (*15–19*). During infection, NSP2 concentrates in viral replication factories, and is essential for virus assembly and genome replication (*20*, *21*). The multifaceted nature of viral RNA chaperones poses various challenges to the dissection of their structure-function relationships, as mutagenesis studies are often limited by intractable pleiotropic effects on viral replication.

The C-terminal region (CTR) of NSP2 (residues 295 – 316) has previously been implicated as important for rotavirus replication (*22–24*) and we have recently demonstrated that C-terminally truncated NSP2 (NSP2-ΔC) has significantly reduced RNA annealing activity in vitro (*17*). The CTR consists of an unstructured linker and an ampholytic helix that contains highly conserved positively- and negatively-charged residues (**Figure 1A**). This suggests that the CTR is required for the RNA chaperone activity of NSP2, although its exact functional role(s) has remained unclear.

Here, we used an orthogonal single-molecule fluorescence spectroscopy approach that allowed us to decouple the RNA annealing and RNA unwinding activities of full-length NSP2 and its truncated variant NSP2-ΔC. We show that while NSP2-ΔC exhibits a reduced capacity to promote RNA-RNA interactions, paradoxically it possesses enhanced RNA unwinding activity. To resolve these apparently incongruous observations, we determined cryo-EM structures of NSP2 and an NSP2-ribonucleoprotein (RNP) complex at global resolutions of 3.9 Å and 3.1 Å, respectively. Consistent with previous biochemical studies, our structural work revealed RNA density localized to surface-exposed positively-charged grooves.

Analysis of conformationally heterogeneous RNP complexes with poorly-defined stoichiometries using conventional structural techniques remains highly challenging. By employing complementary structural proteomics tools, we have mapped the RNA-binding surfaces of NSP2, allowing us to directly visualise the conformationally heterogeneous NSP2-RNA contacts that mediate non-specific, high-affinity RNA binding.

Together, our data reveal that, while the CTR does not directly interact with RNA, it contains a conserved acidic patch that is poised towards bound RNA. We demonstrate that the CTR promotes RNA release, while NSP2-ΔC forms a highly stable, kinetically-trapped RNP complex that is not conducive to RNA annealing. Our multi-pronged approach provides a mechanistic basis for RNA release from high-affinity capture by NSP2, which is required for RNA annealing, chaperone dissociation, and ultimately efficient selection and assembly of a complete stoichiometric

set of distinct RNAs. Furthermore, it allows dissection of complex structure-function relationships of multi-functional RNA-binding proteins by decoupling their helix-unwinding and annealing activities.

Results

Conserved NSP2 CTR is required for efficient RNA annealing

The C-terminal region (CTR) of NSP2 consists of a flexible linker (residues 295 - 304) that tethers an α-helix (C-terminal helix, CTH) to $NSP2_{core}$ (i.e. residues 1 – 294, **Figure 1A**). The CTH is ampholytic, containing highly-conserved positively- (Arg307, Lys308) and negatively-charged (Asp306, Asp310, Glu311) residues. To interrogate the role of the CTR in NSP2 function, we generated a C-terminally truncated NSP2 construct (NSP2-ΔC, residues 1 – 294) lacking the entire CTR (**Supplementary Figure S1**). This NSP2-ΔC construct has been previously characterised by others (*25*, *26*) (**Figure 1**).

To visualise the CTR conformation in solution, we determined a cryo-EM 3D reconstruction of full-length NSP2 (henceforth referred to as NSP2) at a global resolution of 3.9 Å (**Figure 1C; Supplementary Figures S2 & S3**). As expected, our cryo-EM-derived model of NSP2 was highly similar to previously solved crystal structures of NSP2 (the overall RMSD between equivalent Cα atoms of the refined model presented here and PDB 1L9V is 1.124 Å) (*25–27*). Within our density map, the C-terminal helix (CTH) exhibited well-resolved density (local resolutions ranging between 3.6 – 4.0 Å (**Supplementary Figure S2**)). Due to intrinsic flexibility, the linker region was poorly resolved (**Supplementary Figure S4**). We also confirmed that NSP2-ΔC remains octameric by determining a negative-stain EM 3D reconstruction (**Supplementary Figure S1**).

Next, we investigated the role of CTR in the RNA annealing activity of NSP2 using a fluorescence cross-correlation (FCCS)-based RNA-RNA interaction assay (*17*). We chose RNA transcripts S6 and S11, representing rotavirus gene segments 6 and 11, as these have been previously shown to form stable RNA-RNA contacts in the presence of NSP2 (*17*). In brief, fluorescently-labelled rotavirus ssRNA genome segments (S6 and S11) were co-incubated in the absence or presence of either NSP2 or NSP2-ΔC. Ensuing intermolecular interactions were then quantitated in solution by measuring the cross-correlation function (CCF) amplitudes (**Figure 2A**). While a zero CCF amplitude is indicative of non-interacting RNAs, increasing yields of intermolecular interactions result in proportionally higher, non-zero CCF amplitudes (*28*).

Co-incubation of S6 and S11 transcripts alone results in a near zero CCF amplitude indicating that they do not spontaneously interact (**Figure 2B and Supplementary Figure S5**). In contrast, addition of NSP2 to an equimolar mixture of these two RNAs produced a high CCF amplitude, indicative of intermolecular RNA duplex formation (**Figure 2B**). This is in agreement with the role of NSP2 as an RNA chaperone, facilitating the remodelling and annealing of structured RNAs (*17*).

Co-incubation of S6 and S11 in the presence of NSP2–ΔC results in a reduced CCF amplitude (**Figure 2B**), indicating that NSP2-ΔC has a reduced RNA annealing activity relative to full-length NSP2. This is in agreement with our previous observation that NSP2-ΔC has reduced capacity to promote interactions between RV segments S5 and S11 (*17*). In addition, our combined data confirms that CTR plays a role in the RNA chaperone function of NSP2 irrespective of the RNA substrates chosen.

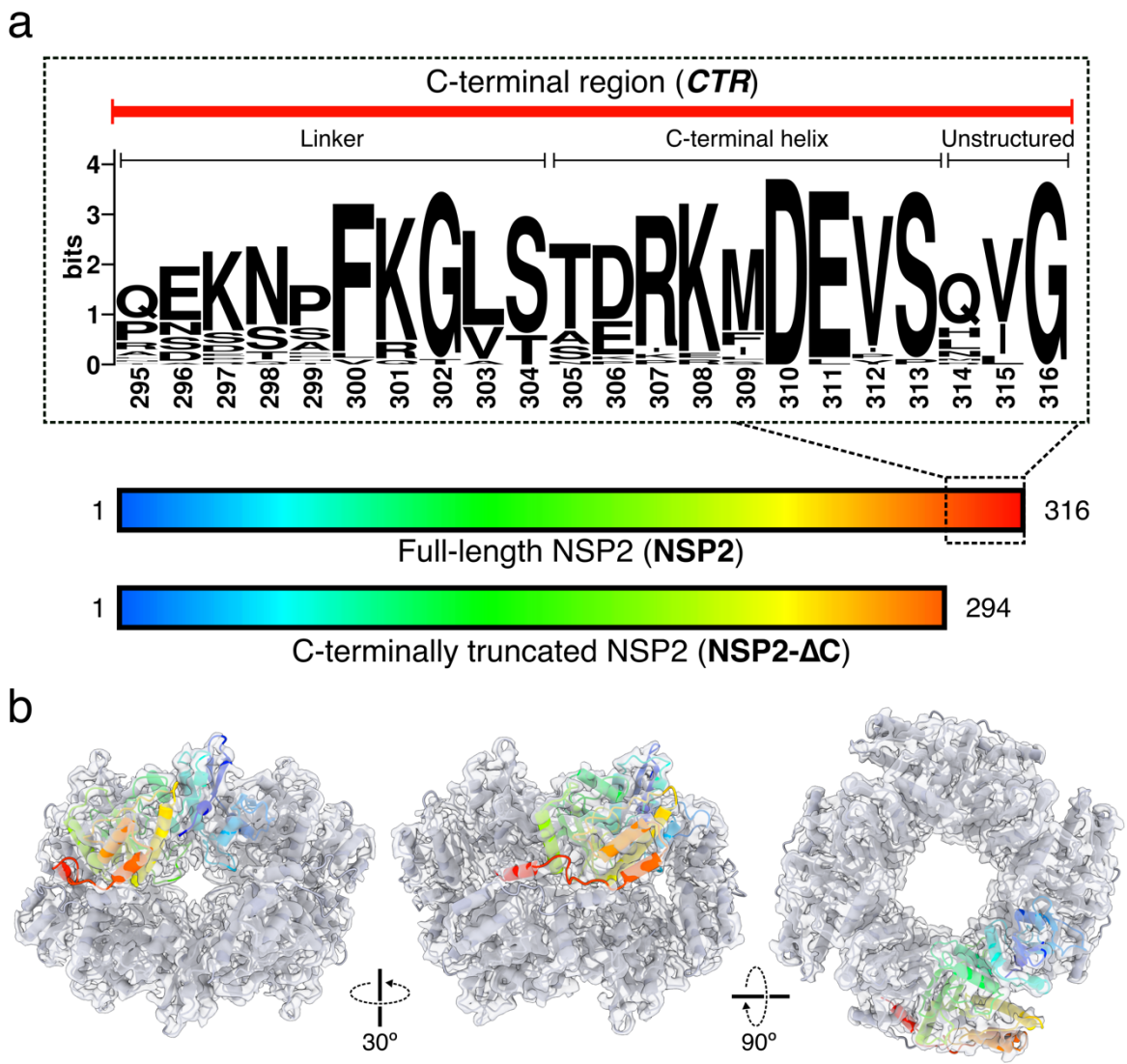


Figure 1. Structure and conservation of NSP2 CTR. A: Constructs of full-length NSP2 (NSP2) and C-terminally truncated NSP2 (NSP2-ΔC, residues 1 – 294) used in this study. An expanded, annotated sequence logo of the NSP2 CTR is shown, which consists of an unstructured, flexible linker region (residues 295 – 304) and a single alpha helix (C-terminal helix, CTH, residues 305 – 313). Downstream residues (314 – 316) are non-essential for viral replication. **B**: A 3.9 Å resolution cryo-EM reconstruction of the octameric NSP2 apoprotein (grey transparent surface) with associated model (cartoon). A single monomer is highlighted and colour coded according to sequence position shown in panel A.

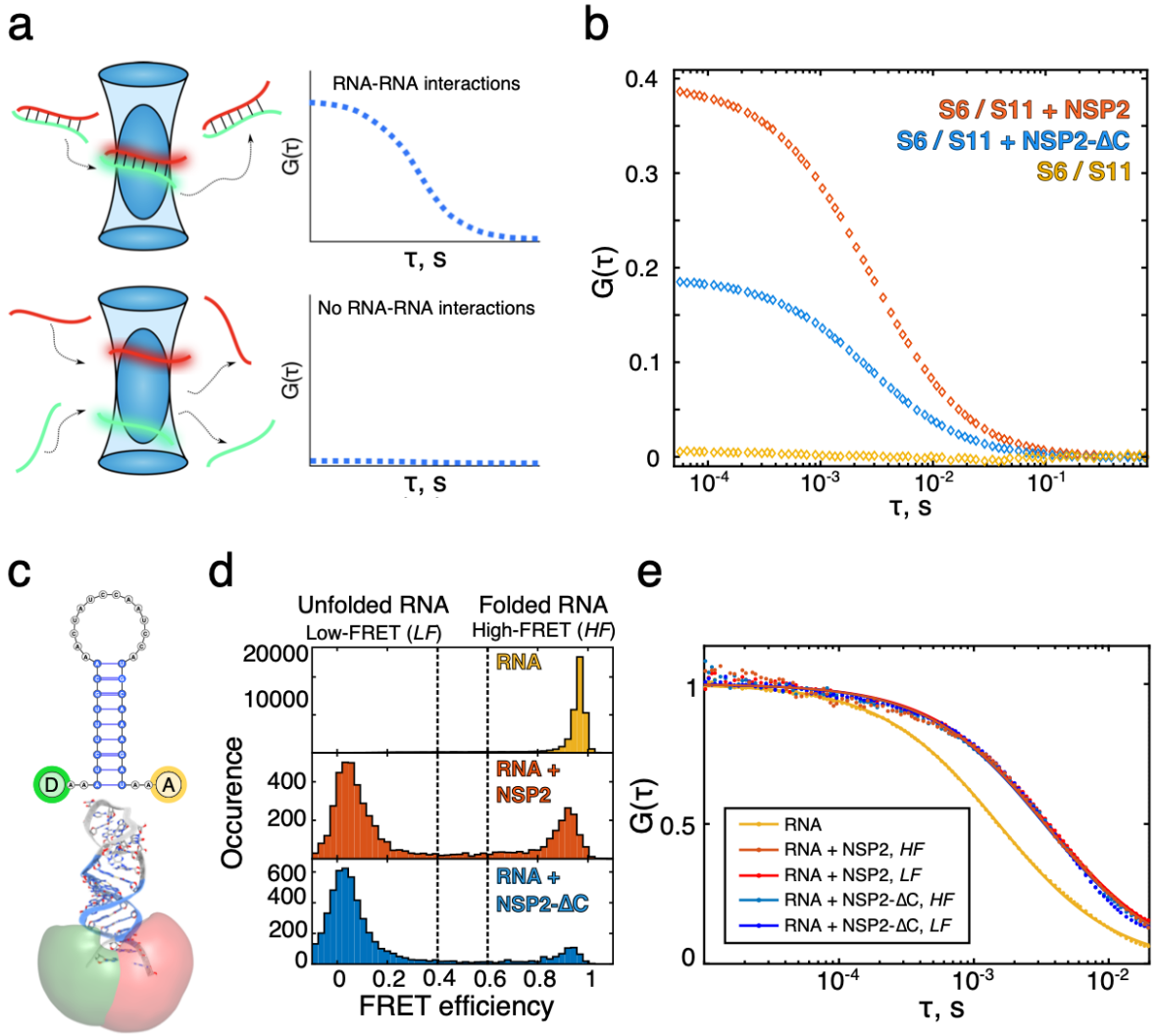


Figure 2. The NSP2 CTR is required for RNA-RNA matchmaking and RNA unwinding. A: Schematics of fluorescence cross-correlation spectroscopy (FCCS) experiment to probe RNA-RNA interactions between fluorescently labelled transcripts S6 (green) and S11 (red). Upon strand annealing, transcripts co-diffuse (shown as a duplex within the blue confocal volume). Such interactions result in a non-zero amplitude of the cross-correlation function G(τ), and thus directly report the fraction of interacting RNAs. A CCF amplitude G(τ) = 0 indicates that the two RNA molecules diffuse independently and are not interacting. **B**: Equimolar amounts of S6 and S11 were co-incubated in the absence (yellow) or presence of either NSP2 (orange) or NSP2-ΔC (blue). While S6 and S11 do not spontaneously interact, co-incubation with NSP2 results in a high fraction of stable S6:S11 complexes. In contrast, co-incubation of S6 and S11 with NSP2-ΔC results in 2-fold reduction of the fraction of S6:S11 complexes. **C**: The RNA stem-loop construct used in the smFRET studies of RNA unwinding activity. The FRET donor (D, green) and acceptor (A, red) dye reporters (Atto532 and Atto647N) and their calculated overlapping accessible volumes (green and red, respectively) are shown. **D**: smFRET efficiency histograms of the RNA stem-loop alone (top, yellow) and in the presence of 5 nM NSP2 (middle, red) or 5 nM NSP2-ΔC (bottom, blue). **E**: A species-selective correlation analysis was performed on the high FRET (*HF*) and low FRET (*LF*) species of RNA stem-loops bound to NSP2 (in orange) and NSP2-ΔC (in blue). For comparison, we also did a species FCS analysis of the free RNA data. All FCS analyses were performed on the smFRET data shown in panel D.

The NSP2 CTR reduces the RNA unwinding activity but does not directly interact with RNA

As the ability of NSP2 to unfold and remodel RNA structures is a prerequisite for its RNA annealing activity (*13*), we next investigated the role of the CTR in RNA helix destabilisation. We used single-molecule Förster Resonance Energy Transfer (smFRET) to directly compare the abilities of NSP2 and NSP2-ΔC to unwind an RNA stem-loop labelled at the 5' and 3' termini with donor and acceptor dyes (Atto532 and Atto647N) (**Figure 2C**).

In the absence of either protein, the stem-loop alone adopts a folded conformation, resulting in a single, high-FRET population (E_{FRET} = ~0.95) (**Figure 2D**). Incubation with NSP2 produces two distinct FRET populations, corresponding to fully folded d(E_{FRET} = ~0.95) and unfolded (E_{FRET} = ~0.05) RNA states. No intermediate FRET populations (corresponding to partially-unwound stem-loop conformations) were observed, in agreement with previous observations of NSP2-mediated RNA unwinding (*13*).

We then measured the ability of NSP2-ΔC to unwind this RNA stem loop. Surprisingly, in the presence of NSP2-ΔC, the stem-loop was predominantly unfolded (E_{FRET} = ~0.05) (**Figure 2D**). Furthermore, we did not observe differences in binding of either NSP2 or NSP2-ΔC to both folded and unfolded RNA conformations (**Figure 2E**). These data demonstrate that NSP2-ΔC has enhanced RNA unfolding activity compared to its full-length counterpart. This result is somewhat paradoxical: while NSP2-ΔC is more efficient at destabilizing RNA structure (**Figure 2D**), it is approximately half as efficient at promoting the annealing of structured RNAs as NSP2 (**Figure 2B**).

To deduce whether the CTR directly interacts with RNA, we used a combination of structural proteomics techniques (**Figure 3**). We performed hydrogen-deuterium exchange-mass spectrometry (HDX-MS) experiments to map regions of NSP2 that become protected from deuterium exchange in the presence of RNA, presumably as they are involved in RNA binding and occluded from solvent when bound. We observed significant protection from exchange for peptides that predominantly mapped to ~25 Å-deep grooves present on the surface of NSP2, indicating that this is the major RNA-binding site of NSP2 (**Figure 3A**). Intriguingly, we did not observe any significant change in protection for peptides that spanned the CTR, indicating that the CTR does not directly interact with RNA (**Figure 3A, Supplementary Figure S7**).

We further corroborated the location of RNA-binding sites on NSP2 using UV-crosslinking with RBDmap (*29*, *30*). Consistent with the HDX-MS data, RBDmap identified RNA-linked peptides map to these surface-exposed RNA-binding grooves (**Figure 3B, Supplementary Figure S7**). However, we again did not observe any RNA-linked peptides corresponding to the CTR.

Collectively, these results reinforce the notion that the CTR is involved in the RNA chaperone activities of NSP2. Our data indicates that although the CTR does not directly interact with RNA, it is a determinant of both the RNA unwinding and annealing activities of NSP2.

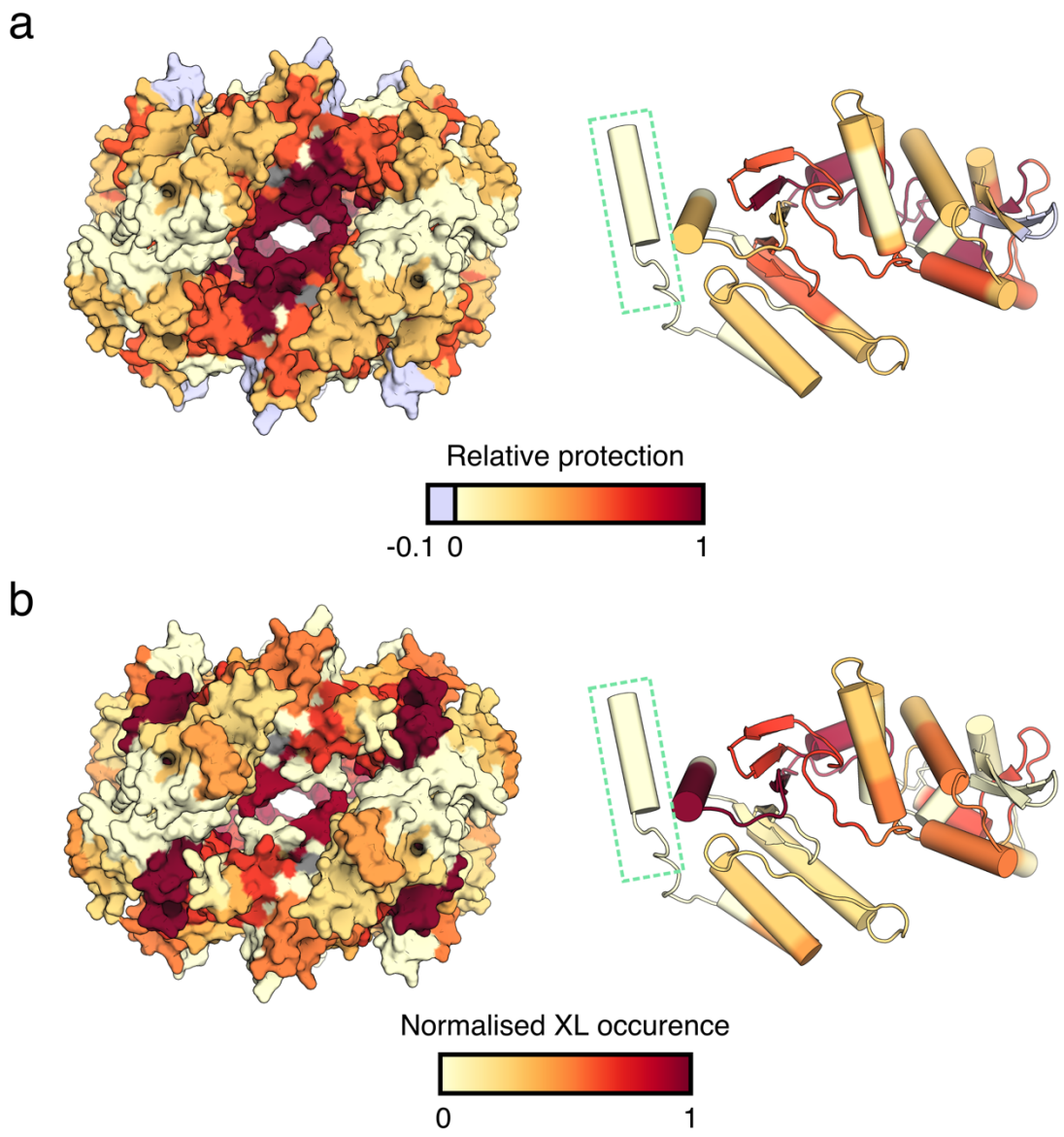


Figure 3. The CTR does not interact with RNA. A: A differential hydrogen-deuterium exchange (HDX) map coloured onto the NSP2 octamer surface (left) and monomer structure (right). Multiple regions of the NSP2 are protected in the presence RNA. Note that, upon RNA binding, the protection rates of the CTR (green box) are not changed. **B**: Normalised occurrence of RNA-interacting peptides determined using UV crosslinking (identified by RBDmap) mapped onto the surface of NSP2 octamer (left) and its monomer (right). Structures are coloured according to crosslink occurrence. No RNA:peptide cross-links are mapped onto the CTR (green box).

Cryo-EM visualization of NSP2-RNA interactions

To understand the molecular basis for the RNA binding by NSP2, we determined a cryo-EM reconstruction of an NSP2 ribonucleoprotein (RNP) complex at a global resolution of 3.1 Å (**Figure 4A, Supplementary Figure S8**). While the cryo-EM density corresponding to NSP2 was well resolved, there was no density that could be attributed to the RNA in the high-resolution post-processed NSP2-RNP map (**Figure 4A**). This is likely due to the heterogeneity and intrinsic flexibility of the NSP2-bound unstructured single-stranded RNA. Despite this, a novel density feature appeared in maps that were low-pass filtered (LPF) to 5 Å resolution (**Figure 4B**) that localised to the RNA binding sites identified by HDX-MS and RBDmap (**Figure 3**) . Notably, such a density feature was not present in 5 Å-LPF NSP2 apoprotein maps (**Figure 4C**). We therefore attribute this density to NSP2-bound RNA.

To improve visualization of the RNA density, we performed focused classification using C4 symmetry-expanded data with a mask applied to a single RNA-binding face of NSP2 (*31–33*). The resulting 3D reconstructions readily classified into four dominant populations, three of which had poor RNA occupancy (each class with 26% of the input particles), while a single 3D class average (22% of input particles) exhibited improved RNA density (**Supplementary Figure S8**). Due to the reasons outlined above, the diffuse nature of this density prevented us from modelling the ssRNA into the structure. However, we were able to visualise residue-specific NSP2-RNA contacts (**Figure 4D**).

We built an atomic model of NSP2 into the sharpened map and then computed a difference map between NSP2 and the RNA-occupied, focused map in order to visualise the NSP2-RNA contacts. Significant positive density was localised in the basic groove of NSP2 (**Figure 4D & E**), consistent with the binding site identified through HDX and RBDmap (**Figure 3**). We observed interactions between positively-charged residues, most notably Arg68 (**Figure 4D**). Adjacent to this contact are Lys58, Lys59, and Arg60, of which Lys59 and Arg60 are directly oriented towards the RNA density (**Figure 4D**, green box). The importance of these residues for RNA capture by NSP2 is strongly supported by previous biochemical studies that identified a number of solvent-exposed lysine and arginine residues (Lys37, Lys38, Lys58, Lys59, Arg60, Arg68) that span the periphery of the NSP2 octamer (**Supplementary Figure S9**) and contribute to RNA binding (*34*).

Furthermore, the identified residues are localised to an unstructured loop within the RNA-binding groove, allowing promiscuous and flexible accommodation of alternative RNA structures with near-identical affinities by NSP2, consistent with previous reports (*13*, *16*). Together with our HDX and RBDmap results, our cryo-EM reconstruction reveals a number of electrostatic contacts that provide a plausible molecular basis for non-specific NSP2-RNA interactions (**Figure 4D**, green box). In addition, our EM reconstruction has revealed a number of other residues (Arg240, Lys286, Phe290) that likely contribute to RNA binding, also identified by HDX and RBDmap (**Supplementary Figures S7 & S9**). These residues may also participate in non-specific RNA contacts via electrostatic interactions, hydrogen bonding and π-π stacking, consistent with a significant non-electrostatic contribution to the overall free energy of RNA binding to NSP2.

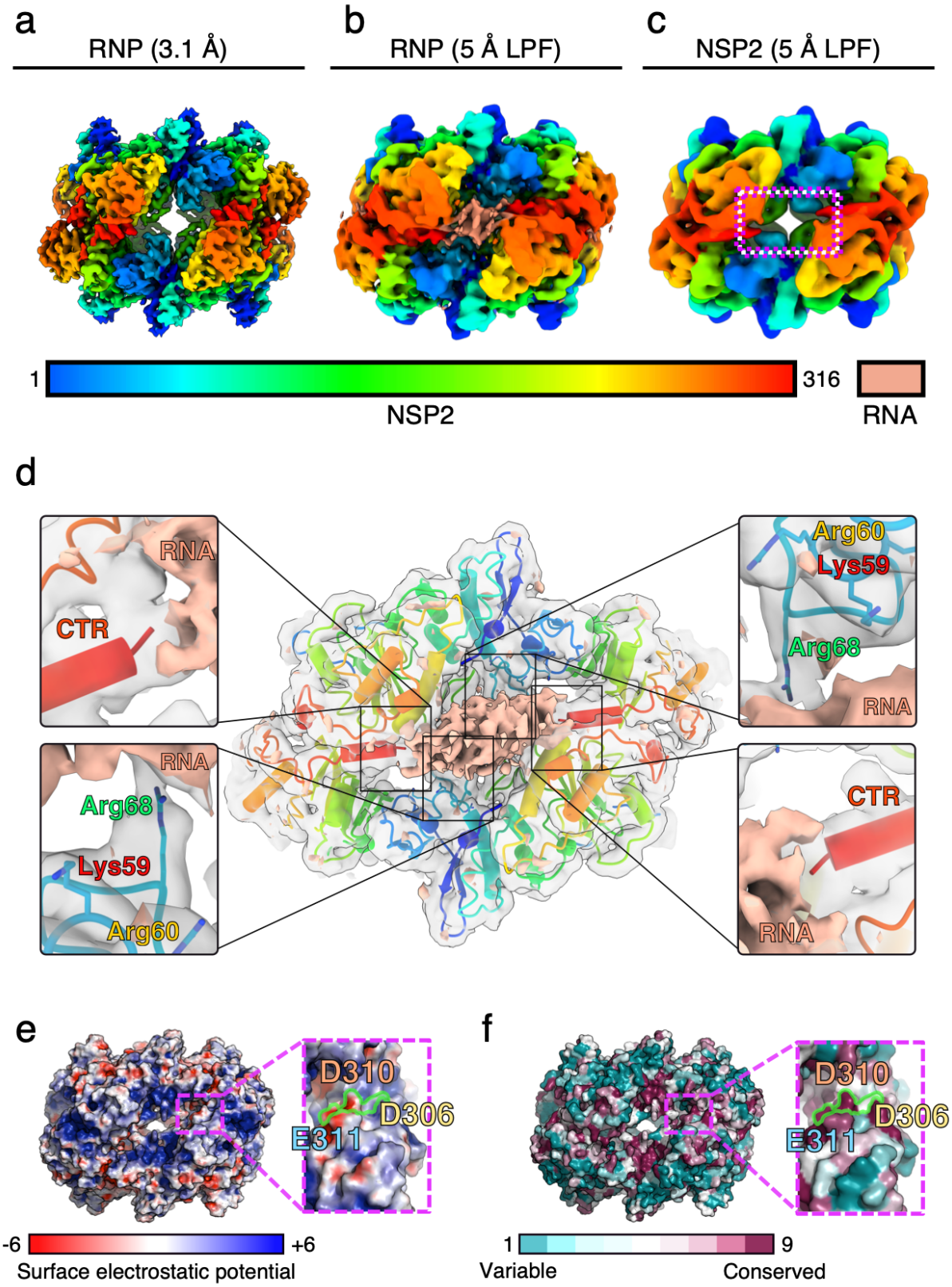


Figure 4. Cryo-EM structure of the NSP2-RNP complex. A: A post-processed 3.1 Å resolution reconstruction of the NSP2-RNP complex, coloured according to corresponding

sequence position (as shown directly underneath). **B & C**: NSP2-RNP (B) and NSP2 apoprotein (C) cryo-EM maps low-pass filtered (LPF) to 5 Å (coloured as in A). A novel electron density attributed to RNA (peach) in the LPF RNP map (B) is highlighted by the dashed box. Both maps are reconstructed with D4 symmetry. **D**: Direct visualisation of interactions between NSP2 and RNA using C4 symmetry expansion and focused classification. The positive difference density map corresponding to RNA (peach) is overlaid onto the unsharpened NSP2-RNP complex map determined through symmetry expansion and focused classification (grey, transparent density) and atomic model of NSP2 (rainbow cartoon, coloured as described above). **E**: The surface electrostatic potential analysis of NSP2 surface is shown (blue is positively-charged; red is negatively charged). **F**: A ConSurf analysis of NSP2 revealing the highly conserved nature of the CTH acidic patch. The surface corresponding to D306, D310 and D311 is outlined in green in panels E & F

Conserved acidic patch within the CTR promotes RNA dissociation

Within the cryo-EM density map, the CTRs are poised below the RNA, while making limited contacts with the observed RNA density (**Figure 4D**). This suggests that rather than modulating the RNA binding affinity, the CTR may play a role in promoting RNA dissociation from the NSP2. To investigate this, we performed binding kinetics measurements using surface plasmon resonance (SPR) (**Figure 5A & B**). Association rate constants (K_{on}) remain largely consistent across a range of concentrations of both NSP2 and NSP2-ΔC (NSP2-ΔC binds 1.5 ± 0.4-fold faster than NSP2 **Table 3**). However, NSP2-ΔC exhibited 3.2 ± 0.3-fold slower dissociation than NSP2, suggesting a role for CTR in the displacement of bound RNA (**Figure 5A**, **Table 3**).

Close examination of our cryo-EM-derived model revealed an acidic patch within the CTR, (**Figure 4E & F**). This is in contrast to other clusters of surface-exposed acidic residues on NSP2 that show a low degree of conservation (**Supplementary figure S10**). The acidic patch of the CTR is presented directly underneath the density attributed to bound RNA (**Figure 4D**), potentially promoting RNA displacement from the NSP2. Such displacement could be achieved either via direct competition with RNA-binding residues or by providing a negatively-charged environment that accelerates RNA dissociation from NSP2 through charge repulsion.

Therefore, to further investigate RNA displacement from NSP2, we used RNA competition assays (**Figure 5C & D**). We performed titrations of unlabelled RNA into preformed RNP complexes containing fluorescently-labelled RNA to understand the differences in RNA exchange and chaperone recycling between NSP2 and NSP2-ΔC. Using fluorescence anisotropy, we estimated the degree of competition as the concentration of competitor RNA required to displace 50% of pre-bound RNA from either NSP2 or NSP2-ΔC complexes (IC_{50}). We determined IC_{50} values of 208 ± 11 nM and 890 ± 160 nM for NSP2 and NSP2-ΔC, respectively, confirming that NSP2–ΔC undergoes ~4-fold reduced RNA exchange, consistent with its ~3-fold slower rate of dissociation from RNA (**Figures 5A & B**).

We then investigated whether the CTR promotes RNA dissociation from NSP2 through directly competing with RNA for binding to basic, RNA-binding residues on the $NSP2_{core}$. To achieve this, we measured RNA binding by NSP2-ΔC in the presence of saturating amounts of the CTR peptide matching the sequence of the NSP2 CTR. No dissociation of RNA from the NSP2-ΔC was observed in the presence of 20-fold molar excess of the CTR peptide over NSP2-ΔC (**Figure 5D**). Furthermore, no RNA

binding was observed upon incubation with 10 μM CTR peptide (i.e. 400-fold excess), indicating that the CTR does not bind RNA. This suggests that, while the CTR is required for RNA displacement from NSP2, this does not occur through direct competition.

We analysed our atomic model of NSP2 to evaluate the distances between acidic residues within the CTR and the basic, RNA-binding residues localised to flexible loops within the RNA-binding grooves (**Figure 5E, Supplementary Figure S11**). The distances (~10 – 30 Å) between acidic residues within the CTR and the RNA-interacting residues are incongruent with a direct competition model. While Arg68 was demonstrated to directly interact with RNA (**Figure 4D**), it is 18 Å away from acidic residues within the CTR. This further demonstrates that, while CTR promotes dissociation of RNA from NSP2, it does not do so through direct competition for $NSP2_{core}$ binding (**Figure 5E**). Collectively, our data suggest that conserved acidic patches within the CTR promote dissociation of bound RNA from NSP2 via charge repulsion.

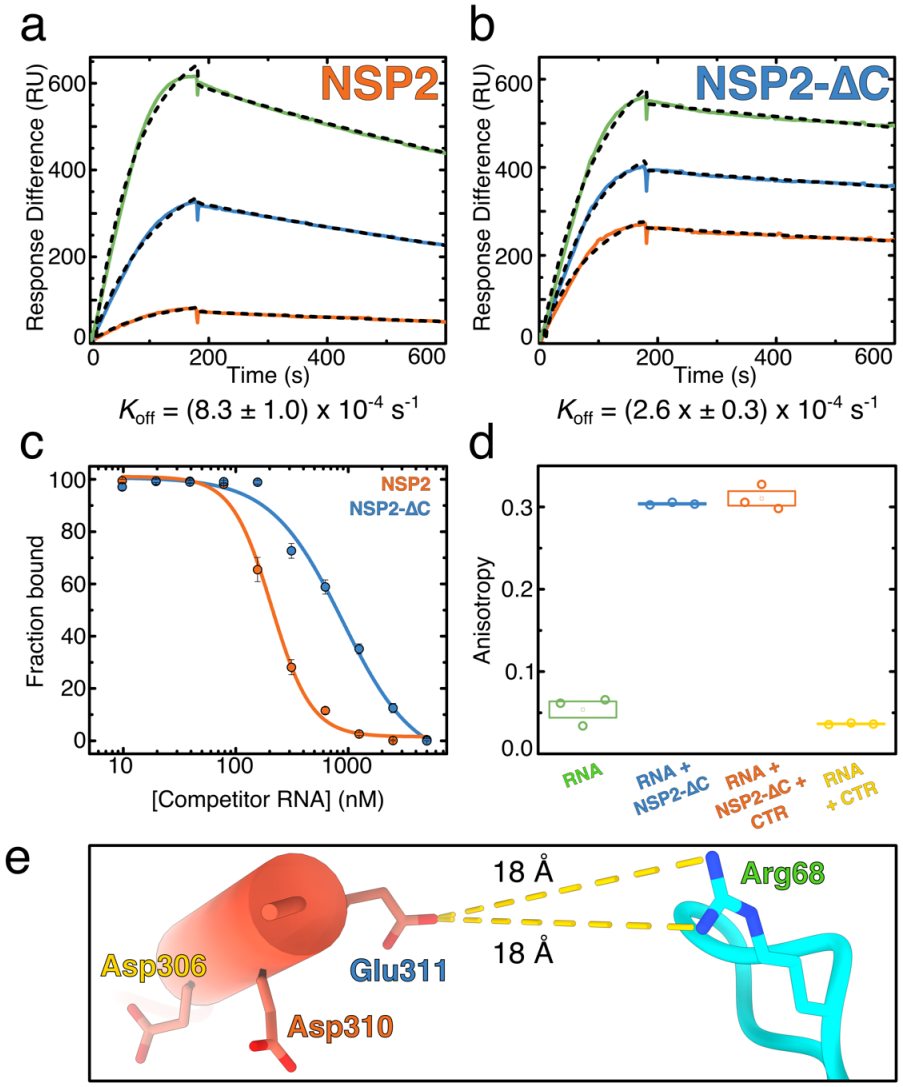


Figure 5. The CTD promotes RNA dissociation non-competitively. A & B: SPR sensograms of NSP2 (A) and NSP2-ΔC (B) binding to RNA. Although NSP2-ΔC binds RNA with ~6-fold higher affinity, this is due to a modest (1.5-fold) increase in K_{on}, and a larger (3.2-fold) decrease in K_{off}. **C**: RNA competition assay. The fractional binding of fluorescently labelled RNA was determined using fluorescence anisotropy. Labelled RNA (10 nM) fully bound to NSP2 (orange) or NSP2-ΔC (blue) was Titrated with unlabelled RNA of identical sequence to compete for NSP2 binding against labelled RNA. The IC_{50} values for NSP2 and NSP2-ΔC were 208 ± 11 nM and 890 ± 160 nM, respectively. The NSP2:RNA complex undergoes strand exchange more readily than the NSP2-ΔC:RNA complex. **D**: We used fluorescence anisotropy experiments to investigate RNA binding to NSP2. Ten μM CTR peptide was added to preformed NSP2-ΔC:RNA complexes. The added CTR peptide did not displace the bound RNA. Hence, a peptide corresponding to the NSP2 CTR does not compete with RNA for binding to NSP2-ΔC. **E**: Distances between acidic residues within CTR and Arg68 interacting with RNA. Note the nearest side chain of Glu311 which is 18A away from Arg68.

Discussion

Long RNAs adopt an ensemble of diverse stable structures that limit spontaneous RNA-RNA interactions through the sequestration of sequences required for intermolecular base pairing (*17*, *35–38*). This necessitates the action of RNA chaperone proteins to bind and refold RNA structures in order to promote RNA annealing between complementary sequences (*39–41*).

In order to function as an RNA chaperone, NSP2 must capture, unwind, anneal and release complementary RNA sequences (*15*, *17*, *42*). Previous structural studies have provided static snapshots of crystallographically-averaged NSP2-RNA complexes (*26*, *27*, *34*). However, due to the highly dynamic nature of the protein-RNA interactions required for its RNA chaperone activity, they have only revealed limited insights into the molecular mechanisms of NSP2. Our recent work (13, 17, 18) indicates that for NSP2-RNP complexes such heterogeneity arises from poorly defined protein-RNA stoichiometries and the ability of bound RNA to adopt multiple configurations and orientations. To overcome these challenges, here we used a combination of single molecule fluorescence, cryo-EM, structural proteomics and biophysical assays to decipher the mechanism of NSP2 chaperone function.

Previous work suggests that the C-terminal region of NSP2 is essential for rotavirus replication (*23*). Using single molecule fluorescence assays, here we have shown that the CTR of NSP2 is important for promoting RNA-RNA interactions. However, we only identified interactions between RNA and basic residues located in flexible loops within the RNA binding groove of NSP2, but not the CTR. Similar RNA recognition mechanisms have been reported in other RNA chaperones including *E. coli* StpA and HIV-1 NC (*43*, *44*). Collectively, these results highlight the role of the CTR in NSP2 RNA chaperone activity but not RNA binding.

Mechanism of the CTR-assisted RNA displacement and its role in RNA matchmaking

FCS analysis of high- and low-FRET RNA species confirms that full length NSP2 binds to both the unfolded and folded RNA conformations, priming RNAs for efficient RNA annealing (**Figure 3E, Supplementary Figure S6**). Single molecule fluorescence and binding kinetics experiments indicate that removal of the CTR does not perturb RNA binding but slows RNA release (~3.2-fold increase in k_{off}). Moreover, CTR removal results in ~2.4-fold increase in RNA-unwinding activity of NSP2-ΔC, as well as a ~2-fold decrease in its RNA annealing activity. Additionally, smFRET data reveal that binding to NSP2-ΔC energetically favours low-FRET (unfolded) RNA conformations, resulting in remodelling of structured RNAs (**Figure 3D**). The resulting increased stability of NSP2-ΔC-RNA complexes precludes efficient RNA annealing, yielding kinetically trapped RNP complex intermediates.

Combined, our data strongly supports a model whereby rotavirus NSP2 binds to RNA with high affinity, resulting in RNA structure destabilisation (**Figure 6A**). By binding to multiple RNAs concurrently via surface-exposed grooves (**Figure 6A, cyan**) (*13*, *17*, *27*), it acts as a matchmaker of complementary sequences, promoting intermolecular RNA-RNA interactions. Conserved acidic patches within the ampholytic CTR (**Figure 6A**, red) accelerate RNA displacement from NSP2 via charge repulsion, thus enabling

RNA chaperone recycling and duplex release. Removal of the ampholytic CTRs in the NSP2 variants derived from two distinct rotavirus strains (SA11 and RF, **Supplementary Figure S12**) has a similar outcome on RNA chaperone activity *in vitro* (**Supplementary Figure S5**), suggesting a conserved role of the CTR in NSP2 function. This model is further supported by our observation that the removal of the unstructured region downstream of the ampholytic CTR (**Figure 1A**) does not alter the RNA unwinding activity of NSP2 (**Supplementary Figure S6**). Interestingly, this partial truncation has been previously shown to support viral replication (*25*).

Remarkably, the described principle of CTR-assisted RNA dissociation from the NSP2 is strikingly similar to that of the bacterial RNA chaperone protein Hfq (*45–47*). Hfq possesses an unstructured C-terminal domain (CTD) with an acidic tip that drives RNA displacement from the Hfq_{core} (*48*). Unlike the Hfq CTD, we do not observe competition between CTR and RNA for $NSP2_{core}$ binding (*49*, *50*). Nevertheless, the NSP2 CTR modulates the kinetics and thermodynamics of NSP2-RNP complex formation, accomplishing RNA chaperone recycling. This may represent a conserved mechanistic feature of multimeric RNA chaperones that capture RNA with high affinity and require auto-regulation to assist RNA dissociation in order to promote efficient matchmaking.

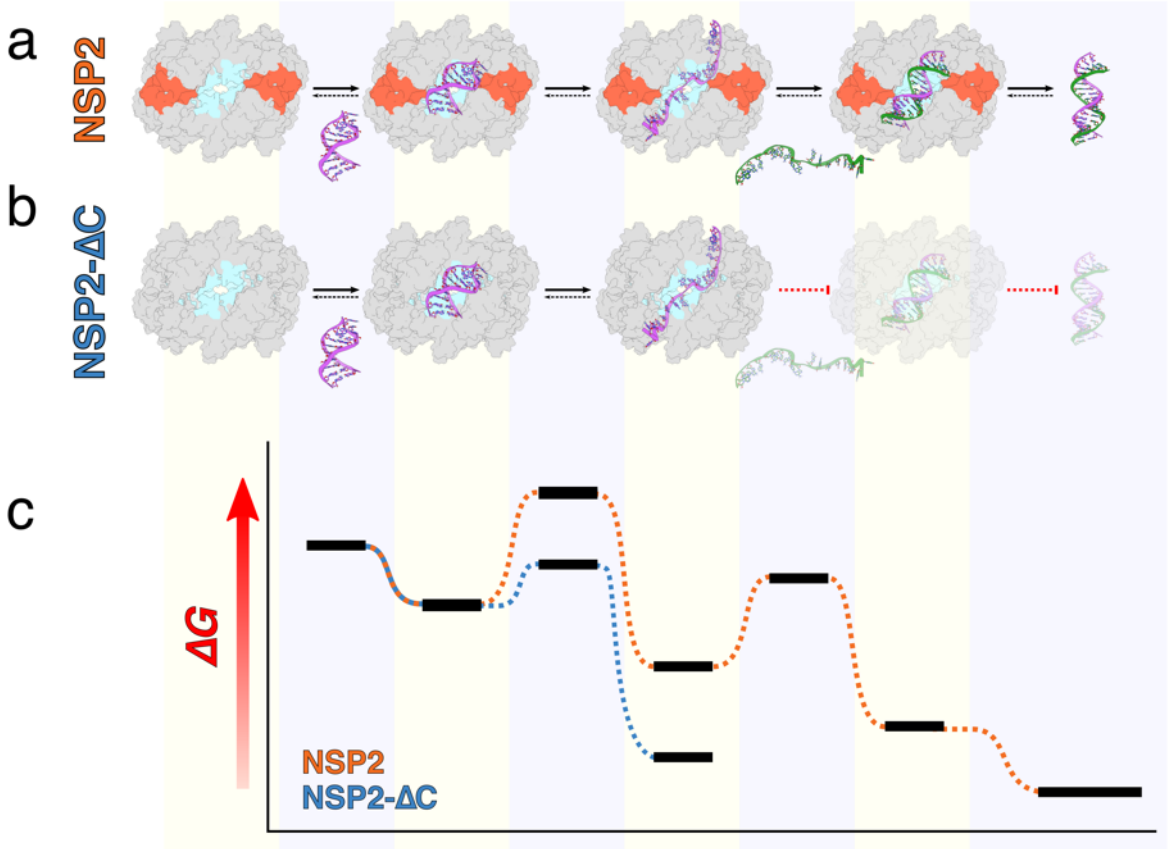


Figure 6. Proposed mechanism of CTD-accelerated RNA dissociation and its requirement for efficient NSP2-mediated RNA-RNA interactions. A: NSP2 captures RNA (purple) via a positively-charged groove (cyan) and promotes RNA unwinding. Binding of a second, complementary RNA strand (green) by NSP2 allows efficient annealing and the proximity to the CTR (burnt orange) promotes dissociation of dsRNA from the NSP2. **B**: In contrast, NSP2-ΔC captures and unwinds RNA, forming a highly stable intermediate. The stability of the intermediate state makes displacement of the bound RNA by a complementary RNA segment via annealing thermodynamically unfavourable. **C**: A free energy diagram of NSP2 (orange) and NSP2-ΔC (blue)-mediated RNA annealing. Horizontal black bars correspond to the free energy levels of different RNA states corresponding to the above schematic representations in A & B.

Materials and Methods

Protein and RNA production

NSP2 and NSP2-ΔC were expressed and purified as described previously (*17*). RNAs used in this study are listed in **Supplementary Table S1**. RNA sequences S5, S6 and S11 were produced and labelled using previously-described *in vitro* transcription and labelling protocols (*17*). Unstructured 20mers (labelled and unlabelled), unstructured 10mer (labelled) and biotinylated unstructured 10mer RNAs were purchased from Integrated DNA technologies (IDT). Dual-labelled stem loop RNA was purchased from IBA Life Sciences.

Negative stain electron microscopy and data processing

For negative-stain grid preparation, 4 μl of sample (at various concentrations ranging from 100 – 500 nM) was incubated on glow-discharged (using PELCO easiGlow) carbon-coated Formvar 300-mesh Cu grid (Agar scientific) for 90 seconds prior to blotting, and stained twice with 20 μl 2% uranyl acetate (first stain immediately blotted, the second stain incubated for 20 seconds prior to blotting) and allowed to dry. Micrographs were collected on an FEI Tecnai 12 transmission electron microscope operated at 120 kV and equipped with a Gatan UltraScan 4000 CCD camera operated at a nominal magnification of 30,000 x (giving a 3.74 Å / pixel sampling on the object level). From 23 micrograph images taken with a nominal defocus of -3 μm, 14,740 particles were picked using template-based autopicking within Relion 3. Multiple rounds of 2D and 3D classification resulted in the selection of a subset of 2,864 particles. These particles were used to determine a ~22 Å resolution NSP2-ΔC reconstruction with D4 symmetry applied.

Cryogenic electron microscopy (cryo-EM) and data processing

Cryo-EM was performed exclusively with Quantifoil R.1.2/1.3 holey carbon grids (i.e. regular ~1.2 μm circular holes with a regular spacing of ~1.3 μm), purchased from Quantifoil. All grids were glow discharged in air using GloQube glow discharge system (Quorum) immediately prior to use. All grids were prepared using a Vitribot IV (FEI) at 100% humidity and 4°C, with a blotting time of 6 seconds and a nominal blotting force of 6. Samples were flash-frozen in liquid nitrogen (LN_2)-cooled liquid ethane and immediately transferred to storage dewars under LN_2.

Vitrified samples were imaged at low temperature in-house (Astbury Biostructure Laboratory, University of Leeds), using Thermo Fisher Titan Krios microscopes equipped with either a Falcon III (NSP2 apoprotein) or a Gatan K2 (NSP2-RNP) detector. Data was collected with an acceleration voltage of 300 kV and a nominal magnification of 75,000x, resulting in pixel sizes of 1.065 Å (Falcon III) or 1.07 Å (K2). Data collection parameters are described in **Supplementary Table 2**.

Image processing was carried out using the Relion 3 pipeline (*51*). Movie drift-correction was performed using MOTIONCOR2 (*52*), and the contrast transfer function of each movie was determined using gCTF (*53*). Initial particle autopicking of a subset of 5 – 10 randomly chosen micrographs was performed with the Laplacian-of-Gaussian (LoG) tool within the Autopicking module of Relion3. Particles were extracted and subjected to initial 2D classification in order to identify particles and

assess autopicking success. Following this, the entire dataset was picked using LoG methods, extracted using 256 pixel box size and binned four times (effective box size 64 pixels) and subjected to 2D classification with fast subsets in order to remove false-positive particles that had been erroneously picked. Next, a more rigorous 2D classification was performed (without fast subsets). Particles originating from 2D classes with secondary structural features were selected and used to generate an initial model. Following multiple rounds of 3D and 2D classification, suitable particles were selected for 3D auto-refinement and various symmetry parameters were applied. Following refinement, per-particle CTF and Bayesian Polishing were performed in Relion 3, and 'shiny' particles were re-refined. Post-processing was performed with a soft mask of 15 pixels and the B-factor estimated automatically in Relion 3.

After particle polishing, the NSP2 apoprotein and the NSP2-RNP complex were subjected to further 3D classification into three classes without particle orientations. This yielded three similarly-sized subsets of near-identical particles for the NSP2 apoprotein whose resolution did not improve upon the original, larger dataset following 3D auto-refinement and post-processing. For the NSP2-RNP complex, this gave a single class with 86% of particles, and two other classes with 6% and 8% of the particles. The class with 6% of input particles had surprisingly well-defined protein and RNA densities and was used for 3D auto-refinement and post-processing. This improved the map resolution from 3.5 Å to 3.4 Å with C4 symmetry. A D4 symmetry reconstruction further increased the map resolution from 3.5 Å to 3.1 Å.

For the NSP2 RNP complex, symmetry expansion was performed on a subset of the 635,599 particles used for a C4 symmetry reconstruction using the *relion_particle_symmetry_expand* command, generating four symmetry-related orientations for each particle. A mask covering a single basic groove-face of NSP2 was made using the volume eraser tool in UCSF ChimeraX (*54*) and Relion 3, with a soft edge of 15 pixels (**Supplementary Figure S8**). The symmetry-expanded dataset was then subjected to focussed classification into 10 classes using this mask without particle orientations. Suitable classes (four classes containing >99% of input particles) were selected, and manually examined for putative RNA density. The subset of particles with the strongest RNA density feature were reconstructed without a mask, and subjected to masking (with the mask corresponding to the entire NSP2 octamer rather than a single face) and post-processing as described for reconstructions with D4 symmetry imposed. Sharpened asymmetric and D4 symmetry maps were aligned using the Fit-In-Map tool within UCSF Chimera and had a correlation of 0.9663.

Atomic model building

A previous atomic model of NSP2 (PDB 1L9V) was fit into the cryo-EM densities using ChimeraX (*54*), and subjected to automated flexible fitting and refinement using Namdinator (*55*). The Namdinator model was used for multiple iterative rounds of manual adjustment in Coot (*56*) and real-space refinement in Phenix (*57*). Models for NSP2 apoprotein and NSP2-RNP were validated using MolProbity (*58*) as implemented in Phenix. Protein sequence conservation and multiple sequence alignment (MSA) was performed using the online ConSurf (*59*) server. Output from the ConSurf MSA was used to generate a sequence logo using the WebLogo server

(*60*). Maps and models were visualized using ChimeraX (*54*) and the electrostatic surfaces were determined using the APBS plugin (*61*).

Single-molecule (sm)FRET measurements

SmFRET measurements of freely diffusing dual-labeled RNA stem-loops in the presence and absence of NSP2 were performed on a home-built confocal microscope as described previously (*13*). Briefly, the samples were excited using pulsed interleaved excitation (*62*) at wavelengths of 532 and 640 nm (PicoTA, Toptica and LDH-D-C-640, PicoQuant) with typical laser powers of 100 μW as measured before the 60x water immersion objective (Plan Apo IR 60x/ 1.27 WI Nikon, Düsseldorf, Germany). The fluorescence signal was split between the green and red detection channels using a DualLine Z532/635 beamsplitter (AHF) and the emission spectra filtered using a Brightline 582/75 filters (Semrock) for green detection and HQ700/75 and ET700/75 filters (Chroma) for red detection. Measurements were performed in eight-well chamber slides (Nunc Lab-Tek, VWR) in a buffer composed of 1/3 PBS (45 mM NaCl, 3 mM phosphate, 1 mM KCl), 1 mM Trolox to reduce photobleaching (*63*) and 0.01% (v/v) tween to prevent sticking of the sample to the glass surface. The dual-labeled RNA stem-loop was diluted to 25 pM and incubated with 5 nM NSP2 (either full length or the ΔC mutant). Data were analyzed with the open-source software package PAM (*64*) using the same burst search parameters and correction factors as described in (*13*). To determine species-selective fluorescence correlation functions, we defined two sub-populations based on the FRET efficiency *E*: the low-FRET population with $E < 0.4$ and the high-FRET population with $E > 0.6$. For each burst, the correlation function for acceptor photons after acceptor excitation was calculated including photons within a time window of 20 ms.

Surface plasmon resonance (SPR)

A Biacore 3000 was used to analyse the binding kinetics of NSP2 and NSP2-ΔC to 5'biotinylated-10mer RNA (**Supplementary Table S1**). All experiments were performed in SPR buffer (150 mM NaCl, 25 mM HEPES, pH 7.5, 0.1 % Tween-20). RNAs were immobilized on an SA sensor chip (GE Healthcare) with an analyte R_{max} of ~20 resonance units (RU). Analyte measurements were performed at 25ºC and a flow rate of 40 μL/min. The chip surface was regenerated between protein injections with a 40 μl 0.05% SDS injection. Data were analysed using BIAevaluation 3.1 software (GE Healthcare). The kinetic parameters were derived assuming a binding stoichiometry of 1 : 1.

RNA competition assay

250 nM NSP2 and NSP2-ΔC (RF) were pre-incubated with 10 nM 20mer AlexaFluor488-labelled RNA (**Supplementary Table S1**) in binding buffer (50 mM NaCl, 25 mM HEPES pH 7.5). Fluorescence anisotropy measurements were performed in the presence of various concentrations of unlabelled 20mer RNA in low-volume Greiner 384-well plates. Data were recorded at 25ºC in a PHERAstar Plus multi-detection plate reader (BMG Labtech) equipped with a fluorescence polarization optical module (λ_{ex} = 485 nm; λ_{em} = 520 nm). The data were normalised and binding curves were fitted in Origin 9.0 using a Hill binding curve resulting in R^2 values of 0.997 and 0.991 for NSP2 and NSP2-ΔC respectively.

CTR peptide competition assay

In order to maximise any potential competition between the CTR peptide and RNA, assays were performed under conditions that favoured dissociation of NSP2-ΔC from RNA. The binding assay was performed in PBS buffer. 25 nM AlexaFluor488-labelled RNA was incubated with 20-fold excess NSP2-ΔC (500 nM) (RF strain). After 30 minutes at room temperature (~25ºC), the CTR peptide was added in 20-fold excess of NSP2-ΔC (i.e. 10 μM). To investigate direct CTR-RNA interactions, 25 nM 10mer RNA was also co-incubated with 10 μM CTR peptide. Fluorescence anisotropy measurements of RNA alone, RNA:NSP2-ΔC, RNA:NSP2-ΔC:CTR and RNA:CTR were performed in triplicate as described above for RNA competition assays.

Hydrogen-deuterium exchange mass spectrometry (HDX-MS)

An automated HDX robot (LEAP Technologies, Ft Lauderdale, FL, USA) coupled to an Acquity M-Class LC and HDX manager (Waters, UK) was used for all HDX-MS experiments. Differential HDX-MS of NSP2 was performed using NSP2 (10 μM) or pre-incubated NSP2-RNP complexes (10 μM + 2 μM 20mer RNA, **Supplementary Table S1**). 30 μl of protein-containing solution was added to 135 µL deuterated buffer (10 mM potassium phosphate buffer pD 8.0, 82% D_2O) and incubated at 4 °C for 0.5, 2, 30 or 120 min. After labelling, HDX was quenched by adding 100 µL of quench buffer (10 mM potassium phosphate, 2 M Gdn-HCl, pH 2.2) to 50 µL of the labelling reaction. 50 µL of the quenched sample was passed through immobilised pepsin and aspergillopepsin columns (Affipro, Mratín, Czech Republic) connected in series (20 °C) and the peptides were trapped on a VanGuard Pre-column [Acquity UPLC BEH C18 (1.7 µm, 2.1 mm × 5 mm, Waters, UK)] for 3 min. The peptides were separated using a C18 column (75 µm × 150 mm, Waters, UK) by gradient elution of 0–40% (v/v) acetonitrile (0.1% v/v formic acid) in H_2O (0.3% v/v formic acid) over 7 min at 40 µL min^{-1}. Peptides were detected using a Synapt G2Si mass spectrometer (Waters, UK). The mass spectrometer was operated in $HDMS^E$ mode with the dynamic range extension enabled (data independent analysis (DIA) coupled with IMS separation) were used to separate peptides prior to CID fragmentation in the transfer cell. CID data were used for peptide identification and uptake quantification was performed at the peptide level (as CID results in deuterium scrambling). Data were analysed using PLGS (v3.0.2) and DynamX (*65*) (v3.0.0) software (Waters, UK). Restrictions for peptides in DynamX were as follows: minimum intensity = 1000, minimum products per amino acid = 0.3, max sequence length = 25, max ppm error = 5, file threshold = 3. The software Deuteros (*66*) was used to identify peptides with statistically significant increases/decreases in deuterium uptake (applying a 99 % confidence interval) and to prepare Woods plots.

UV-crosslinking-mass spectrometry with RBDmap

10 μM NSP2 was incubated with 5'-A_{25}-S11 RNA in a final volume of 100 μl. RNP complexes were incubated at room temperature for 30 minutes and applied to a single well of a 24-well plate. This 24-well plate was placed on an aluminium block cooled to 4°C within a plastic container of ice and subjected to 6 rounds of UV irradiation (254 nm, 0.83 J cm^{-2} per round) in a UVP CL-1000 Ultraviolet Crosslinker (Scientifix). Crosslinked RNP complexes were digested by LysC (NEB, #P8109S) (500 ng per crosslinked RNP sample) overnight at room temperature. Enrichment and

identification of cross-linked peptides were performed using the in vitro adaptation of the RBDmap protocol, as described in (*29*). Data analysis was performed using the CrissCrosslinker R script, as described in (*29*).

Circular dichroism (CD)

CD experiments were performed in a Chirascan plus spectrometer (Applied Photophysics). Samples were prepared by dialyzing protein solutions against 10 mM phosphate buffer pH 7.4, 50 mM sodium fluoride. Spectra were recorded over a wavelength range of 190–260 nm with a bandwidth of 1 nm, step size of 1 nm and a path length of 1 mm. An average of three scans were used for the final spectra.

NSP2 (RF) threading and sequence alignment

Alignment of NSP2(SA11) and NSP2(RF) sequences was performed using T-Coffee (*67*). Threading of the NSP2(RF) sequence (based on the SA11 structure, PDB 1L9V) was performed using I-TASSER (*68*).

Authors' contribution: J.B., K.B., E.G., A.C. and A.B. designed and carried out experiments, and analyzed data. J.B., K.B., A.C. and A.B. jointly wrote the manuscript. R.T., D.C.L., C.D. contributed novel analytical tools. J.B. collected and analyzed EM data. A.C. collected and analyzed HDX data. K.B. and A.B. collected and analysed single-molecule fluorescence data. A.B. managed the project. All authors contributed ideas, discussed the results and were involved in writing of the manuscript.

Acknowledgements

Wellcome Trust [103068/Z/13/Z and 213437/Z/18/Z to A.B.]; Biotechnology and Biological Sciences Research Council (BBSRC) White Rose DTP [BB/M011151/1 to J.P.K.B.]; European Regional Development Fund [CZ.02.1.01/0.0/0.0/15_003/0000441 to R.T.]; A.N.C. acknowledges support from a University Academic Fellowship from the University of Leeds. E.H.G. holds a Biomedicine Discovery Scholarship and is an EMBL-Australia PhD student. C.D. is an EMBL-Australia Group Leader and acknowledges support from the ARC (DP190103407) and the NHMRC (APP1162921 & APP1184637).

Deutsche Forschungsgemeinschaft SFB1032 (Project B3) [to D.C.L.] and the Ludwig-Maximilians-Universität, München through the Center for NanoScience (CeNS) and the LMUinnovativ initiative BioImaging Network (BIN) (to D.C.L.).

Funding from the BBSRC (BB/M012573/1) to purchase HDX-MS instrumentation is acknowledged.

Funding for open access charge: Wellcome Trust.

References

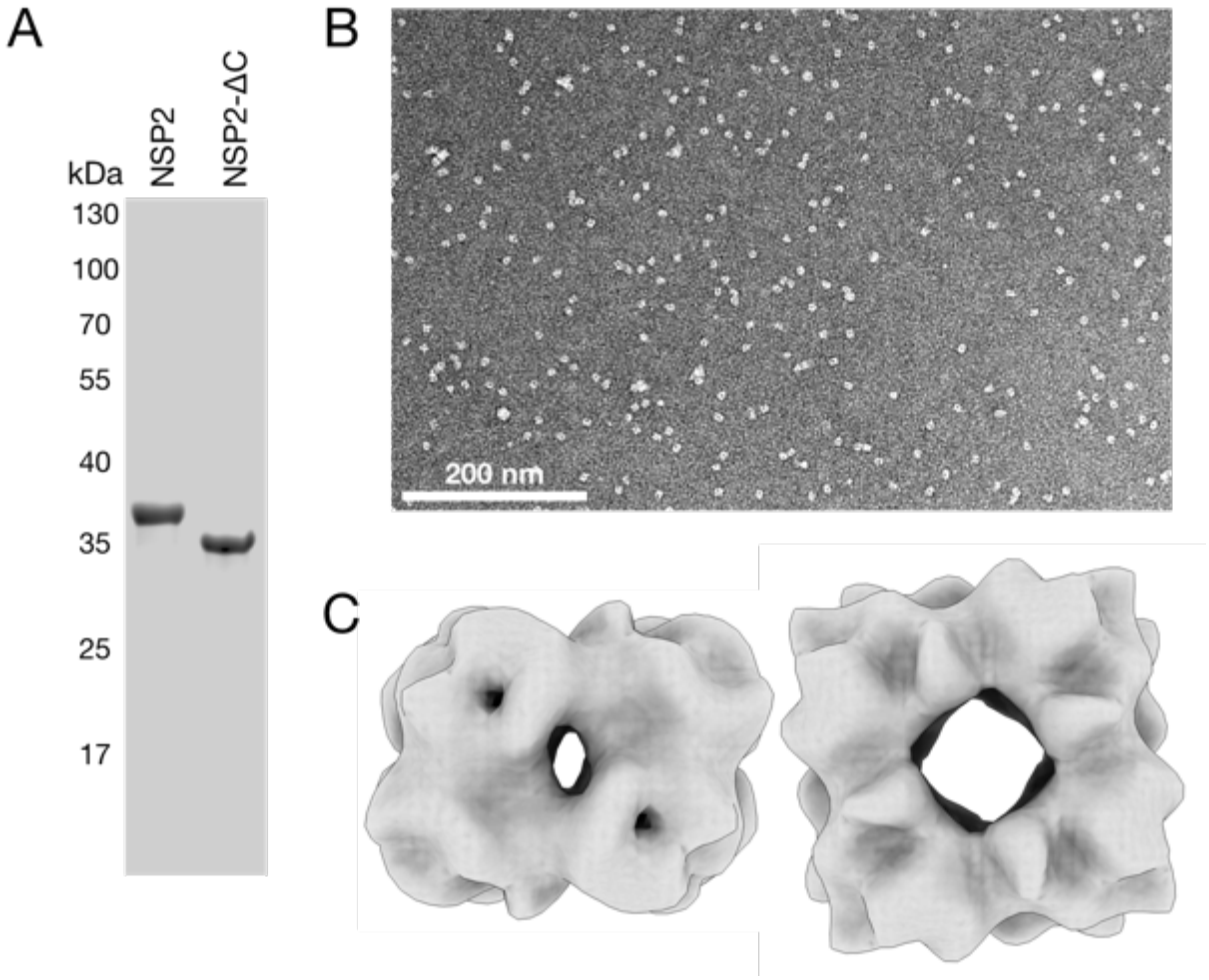
1. Z. A. Jaafar, J. S. Kieft, Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol.* **17** (2019), pp. 110–123.
2. A. Serganov, D. J. Patel, Molecular recognition and function of riboswitches. *Curr. Opin. Struct. Biol.* **22**, 279–286 (2012).
3. J. A. Cruz, E. Westhof, The Dynamic Landscapes of RNA Architecture. *Cell*. **136**, 604–609 (2009).
4. R. T. Batey, R. P. Rambo, J. A. Doudna, Tertiary motifs in RNA structure and folding. *Angew. Chemie - Int. Ed.* **38**, 2326–2343 (1999).
5. B. L. Nicholson, K. A. White, Functional long-range RNA – RNA interactions in positive-strand RNA viruses. *Nat. Publ. Gr.* **12**, 493–504 (2014).
6. J. D. Perlmutter, M. F. Hagan, Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66**, 217–239 (2015).
7. A. Rein, RNA Packaging in HIV. *Trends Microbiol.* **27**, 715–723 (2019).
8. S. McDonald, M. I. Nelson, P. E. Turner, J. T. Patton, Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat. Rev. Microbiol.* **14**, 448–460 (2016).
9. S. McDonald, J. T. Patton, Assortment and packaging of the segmented rotavirus genome. *Trends Microbiol.* **19**, 136–144 (2011).
10. A. Borodavka, U. Desselberger, J. T. Patton, Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. *Curr. Opin. Virol.* **33**, 106–112 (2018).
11. L. R. Newburn, K. A. White, Trans-Acting RNA–RNA Interactions in Segmented RNA Viruses. *Viruses*. **11**, 751 (2019).
12. Z. F. Taraporewala, J. T. Patton, Nonstructural proteins involved in genome packaging and replication of rotaviruses and other members of the Reoviridae. *Virus Res.* **101** (2004), pp. 57–66.
13. J. P. K. Bravo, A. Borodavka, A. Barth, A. N. Calabrese, P. Mojzes, J. J. B. Cockburn, D. C. Lamb, R. Tuma, Stability of local secondary structure determines selectivity of viral RNA chaperones. *Nucleic Acids Res.*, 293191 (2018).
14. Z. F. Taraporewala, J. T. Patton, D. Chen, J. T. Patton, Identification and Characterization of the Helix-Destabilizing Activity of Rotavirus Nonstructural Protein NSP2. *J. Virol.* **75**, 4519–4527 (2001).
15. Z. F. Taraporewala, J. T. Patton, Identification and Characterization of the Helix-Destabilizing Activity of Rotavirus Nonstructural Protein NSP2. *J. Virol.* **75**, 4519–4527 (2001).
16. Z. F. Taraporewala, D. Chen, J. T. Patton, Multimers Formed by the Rotavirus Nonstructural Protein NSP2 Bind to RNA and Have Nucleoside Triphosphatase Activity. *J. Virol.* **73**, 9934–9943 (1999).
17. A. Borodavka, E. C. Dykeman, W. Schrimpf, D. C. Lamb, Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *Elife*. **6**, 1–22 (2017).
18. A. Borodavka, J. Ault, P. G. Stockley, R. Tuma, Evidence that avian reovirus σNS is an RNA chaperone: implications for genome segment assortment. *Nucleic Acids Res.* **43**, 7044–7057 (2015).
19. P. Schuck, Z. F. Taraporewala, P. McPhie, J. T. Patton, Rotavirus Nonstructural Protein NSP2 Self-assembles into Octamers that Undergo Ligand-induced Conformational Changes. *J. Biol. Chem.* **276**, 9679–9687 (2001).
20. G. Papa, L. Venditti, F. Arnoldi, E. M. Schraner, C. Potgieter, A. Borodavka, C. Eichwald, O. R. Burrone, Recombinant rotaviruses rescued by reverse genetics reveal the role of NSP5 hyperphosphorylation in the assembly of viral factories. *J.*

Virol. **94**, 1–23 (2019).
21. R. Mingo, S. Zhang, C. P. Long, L. E. W. Laconte, S. McDonald, Genetic determinants restricting the reassortment of heterologous NSP2 genes into the simian rotavirus SA11 genome. *Sci. Rep.*, 1–12 (2017).
22. L. Hu, S. E. Crawford, J. M. Hyser, M. K. Estes, B. V. V. Prasad, Rotavirus non-structural proteins: Structure and function. *Curr. Opin. Virol.* **2**, 380–388 (2012).
23. J. M. Criglar, R. Anish, L. Hu, S. E. Crawford, B. Sankaran, B. V. V. Prasad, M. K. Estes, Phosphorylation cascade regulates the formation and maturation of rotaviral replication factories. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E12015–E12023 (2018).
24. J. M. Criglar, L. Hu, S. E. Crawford, J. M. Hyser, J. R. Broughman, B. V. V. Prasad, M. K. Estes, A Novel Form of Rotavirus NSP2 and Phosphorylation-Dependent NSP2-NSP5 Interactions Are Associated with Viroplasm Assembly. *J. Virol.* **88**, 786–98 (2014).
25. Z. F. Taraporewala, X. Jiang, R. Vasquez-Del Carpio, H. Jayaram, B. V. V. Prasad, J. T. Patton, Structure-function analysis of rotavirus NSP2 octamer by using a novel complementation system. *J. Virol.* **80**, 7984–7994 (2006).
26. L. Hu, D.-C. Chow, J. T. Patton, T. Palzkill, M. K. Estes, B. V. V. Prasad, Crystallographic Analysis of Rotavirus NSP2-RNA Complex Reveals Specific Recognition of 5' GG Sequence for RTPase Activity. *J. Virol.* **86**, 10547–10557 (2012).
27. H. Jayaram, Z. F. Taraporewala, J. T. Patton, B. V. V. Prasad, Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature*. **417**, 311–5 (2002).
28. E. Haustein, P. Schwille, Single-molecule spectroscopic methods. *Curr. Opin. Struct. Biol.* **14** (2004), pp. 531–540.
29. Q. Zhang, N. J. McKenzie, R. Warneford-Thomson, E. H. Gail, S. F. Flanigan, B. M. Owen, R. Lauman, V. Levina, B. A. Garcia, R. B. Schittenhelm, R. Bonasio, C. Davidovich, RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat. Struct. Mol. Biol.* **26**, 237–247 (2019).
30. A. Castello, B. Fischer, C. K. Frese, R. Horos, A. M. Alleaume, S. Foehr, T. Curk, J. Krijgsveld, M. W. Hentze, Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell*. **63**, 696–710 (2016).
31. M. J. Conley, M. McElwee, L. Azmi, M. Gabrielsen, O. Byron, I. G. Goodfellow, D. Bhella, Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature*. **565** (2019), pp. 377–381.
32. S. H. W. Scheres, *Processing of Structurally Heterogeneous Cryo-EM Data in RELION* (Elsevier Inc., ed. 1, 2016; http://dx.doi.org/10.1016/bs.mie.2016.04.012), vol. 579.
33. D. Lyumkis, Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **294**, 5181–5197 (2019).
34. R. Vasquez-Del Carpio, F. D. Gonzalez-Nilo, G. Riadi, Z. F. Taraporewala, J. T. Patton, Histidine Triad-like Motif of the Rotavirus NSP2 Octamer Mediates both RTPase and NTPase Activities. *J. Mol. Biol.* **362**, 539–554 (2006).
35. D. Thirumalai, S. A. Woodson, Kinetics of Folding of Proteins and RNA. *Acc. Chem. Res.* **29**, 433–439 (1996).
36. M. A. Ditzler, D. Rueda, J. Mo, K. Håkansson, N. G. Walter, A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.* **36**, 7088–7099 (2008).
37. H. F. Noller, RNA Structure: Reading the Ribosome. *Science (80-.).* **309**, 1508–1514 (2005).
38. P. Bieniasz, A. Telesnitsky, *Annu. Rev. Virol.*, in press, doi:10.1146/annurev-virology-

092917-043448.
39. D. Herschlag, RNA chaperones and the folding problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).
40. R. Russell, RNA misfolding and the action of chaperones. *Front. Biosci.* **13**, 1–20 (2008).
41. I. Jarmoskaite, R. Russell, DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdiscip. Rev. RNA*. **2** (2011), pp. 135–152.
42. L. Rajkowitsch, D. Chen, S. Stampfl, K. Semrad, C. Waldsich, O. Mayer, M. F. Jantsch, R. Konrat, U. Bläsi, R. Schroeder, RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **4** (2007), pp. 118–130.
43. O. Mayer, L. Rajkowitsch, C. Lorenz, R. Konrat, R. Schroeder, RNA chaperone activity and RNA-binding properties of the E. coli protein StpA. *Nucleic Acids Res.* **35**, 1257–1269 (2007).
44. A. Belfetmi, L. Zargarian, C. Tisné, D. Sleiman, N. Morellet, E. Lescop, O. Maskri, B. René, Y. Mély, P. Fossé, O. Mauffret, Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein. *Rna*. **22**, 506–517 (2016).
45. J. Vogel, B. F. Luisi, Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9** (2011), pp. 578–589.
46. A. Santiago-Frangos, S. A. Woodson, Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip. Rev. RNA*. **9**, 1–16 (2018).
47. S. A. Woodson, S. Panja, A. Santiago-Frangos, Proteins That Chaperone RNA Regulation. *Microbiol. Spectr.* **113**, E6089–E6096 (2018).
48. A. Santiago-Frangos, K. Kavita, D. J. Schu, S. Gottesman, S. A. Woodson, C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA. *Proc. Natl. Acad. Sci.* **113**, E6089–E6096 (2016).
49. A. Santiago-Frangos, K. S. Fröhlich, J. R. Jeliazkov, E. M. Małecka, G. Marino, J. J. Gray, B. F. Luisi, S. A. Woodson, S. W. Hardwick, Caulobacter crescentus Hfq structure reveals a conserved mechanism of RNA annealing regulation. *Proc. Natl. Acad. Sci. U. S. A.*, 1–10 (2019).
50. A. Santiago-Frangos, J. R. Jeliazkov, J. J. Gray, S. A. Woodson, Acidic C-terminal domains autoregulate the RNA chaperone Hfq. *Elife*. **6**, 1–25 (2017).
51. J. Zivanov, T. Nakane, B. Forsberg, D. Kimanius, W. J. H. J. Hagen, E. Lindahl, S. H. W. Scheres, E. H. Egelman, J. Zivanov, T. Nakane, B. O. rn Forsberg, D. Kimanius, W. J. H. J. Hagen, E. Lindahl, S. H. W. Scheres, RELION-3: new tools for automated high-resolution cryo-EM structure determination. *Elife*, 421123 (2018).
52. S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*. **14**, 331–332 (2017).
53. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
54. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
55. R. T. Kidmose, J. Juhl, P. Nissen, T. Boesen, J. L. Karlsen, B. P. Pedersen, Namdinator - Automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ*. **6**, 526–531 (2019).
56. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
57. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and

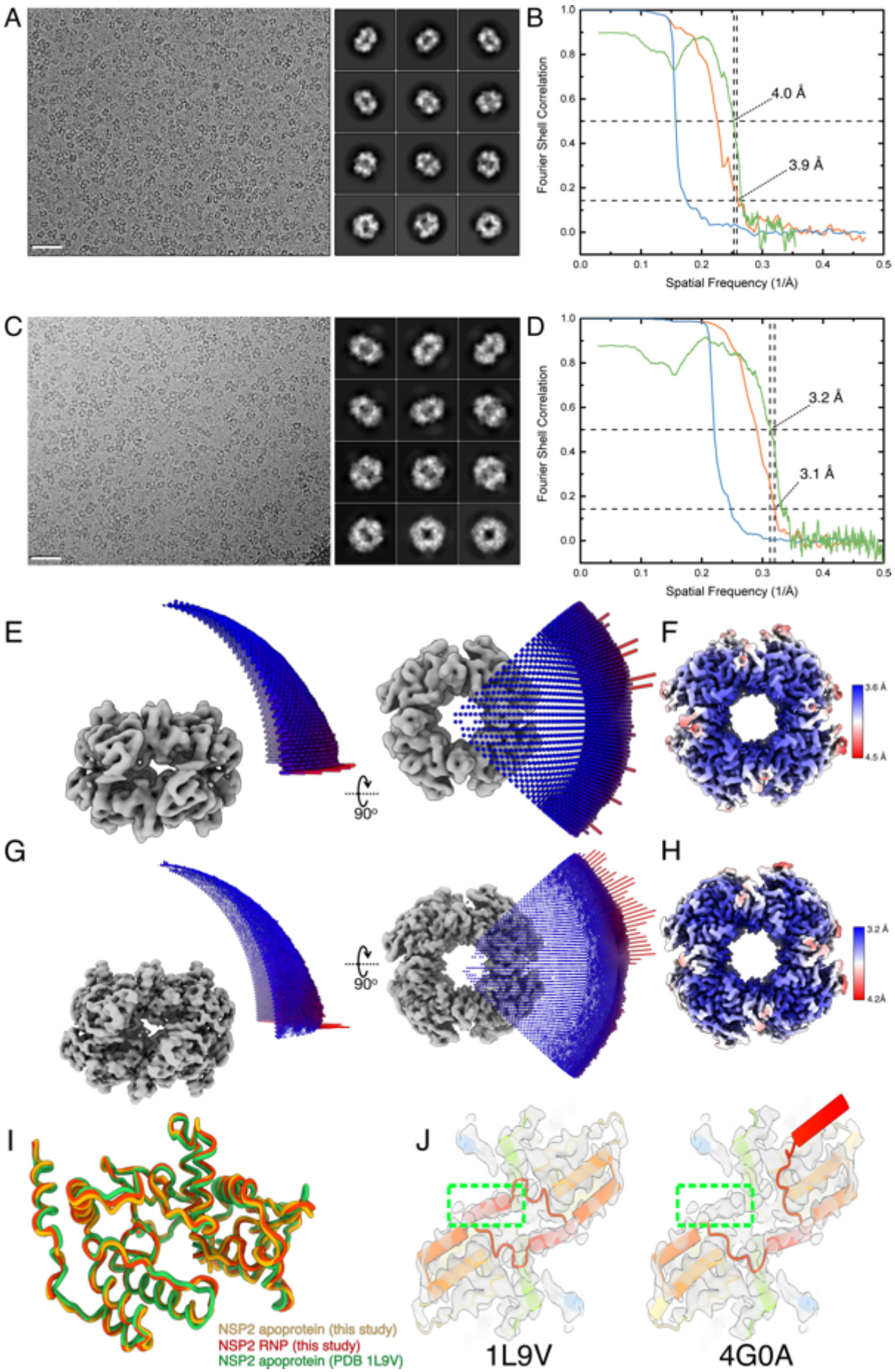
crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).
58. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
59. H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
60. G. Crooks, G. Hon, J. Chandonia, S. Brenner, WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188–1190 (2004).
61. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041 (2001).
62. V. Kudryavtsev, M. Sikor, S. Kalinin, D. Mokranjac, C. A. M. Seidel, D. C. Lamb, Combining MFD and PIE for accurate single-pair förster resonance energy transfer measurements. *ChemPhysChem*. **13**, 1060–1078 (2012).
63. T. Cordes, J. Vogelsang, P. Tinnefeld, On the mechanism of trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.* **131**, 5018–5019 (2009).
64. W. Schrimpf, A. Barth, J. Hendrix, D. C. Lamb, Computational Tool PAM : A Framework for Integrated Analysis of Imaging , Single-Molecule , and Ensemble Fluorescence Data. *Biophysj*. **114**, 1518–1528 (2018).
65. A. Cryar, K. Groves, M. Quaglia, Online Hydrogen-Deuterium Exchange Traveling Wave Ion. *J. Am. Soc. Mass Spectrom.* **28**, 1192–1202 (2017).
66. A. M. C. Lau, Z. Ahdash, C. Martens, A. Politis, Deuteros: Software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry. *Bioinformatics*. **35**, 3171–3173 (2019).
67. P. Di Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J. M. Chang, J. F. Taly, C. Notredame, T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, 13–17 (2011).
68. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).

Supplementary Information

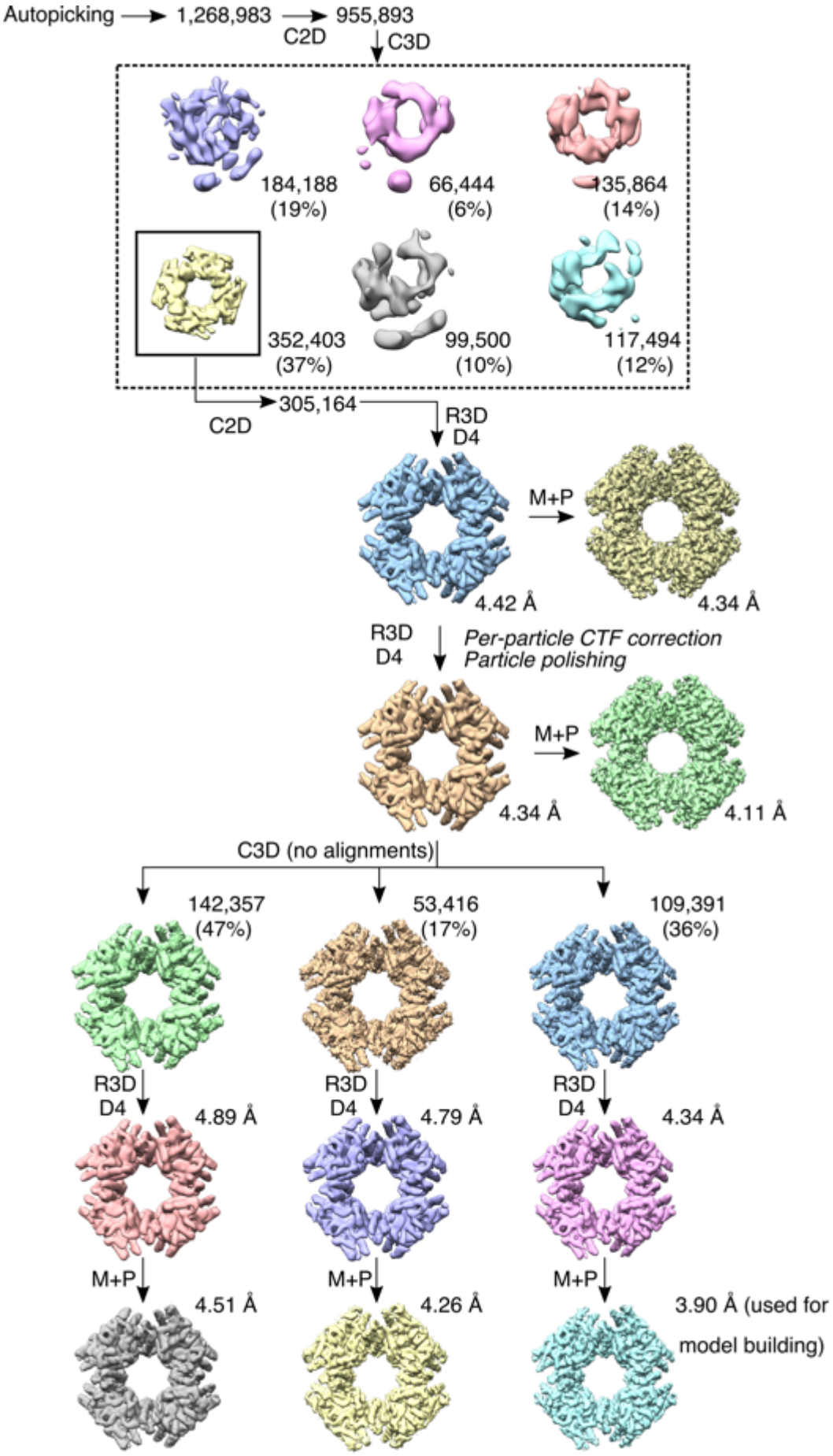


Supplementary Figure S1. NSP2–ΔC is octameric.

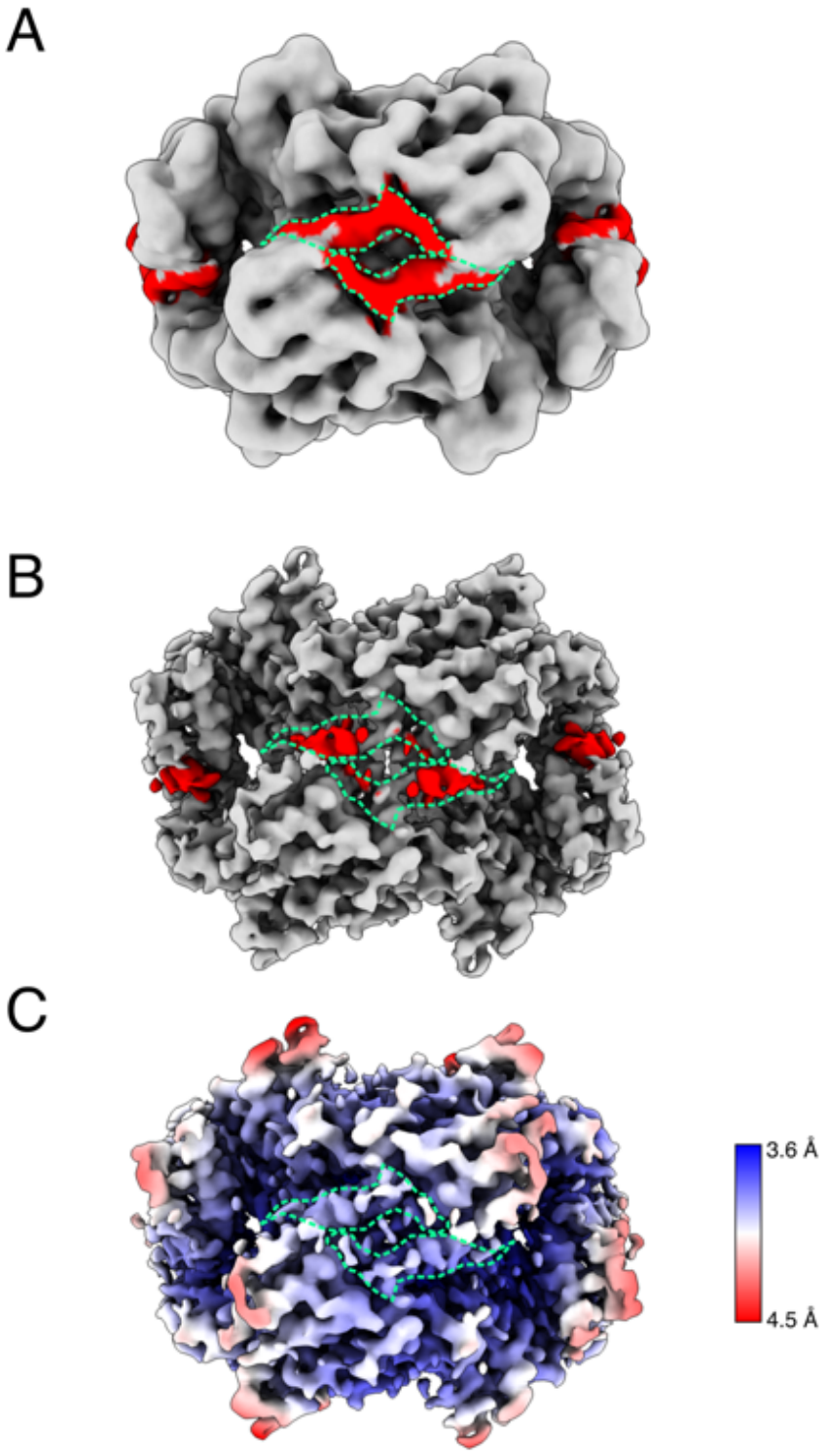
A: SDS-PAGE of purified NSP2 and NSP2-ΔC. **B:** Representative negative stain EM micrograph of NSP2-ΔC **C:** 3D reconstruction of NSP2-ΔC.


A
B
Fourier Shell Correlation
4.0 Å
3.9 Å
Spatial Frequency (1/Å)
C
D
Fourier Shell Correlation
3.2 Å
3.1 Å
Spatial Frequency (1/Å)
E
90°
F
3.6 Å
4.5 Å
G
90°
H
3.2 Å
4.2Å
I
NSP2 apoprotein (this study)
NSP2 RNP (this study)
NSP2 apoprotein (PDB 1L9V)
J
1L9V
4G0A

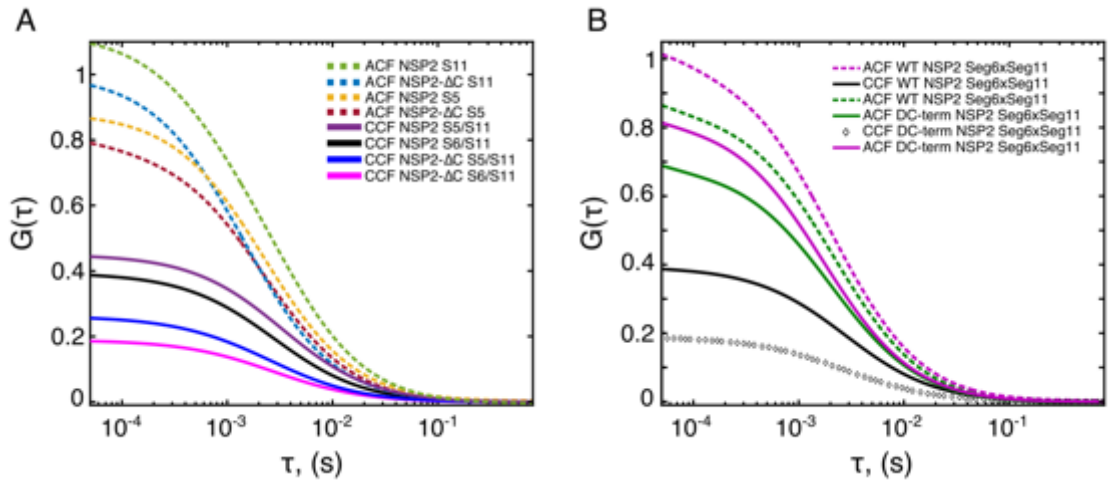
Supplementary Figure S2. Cryo-EM **structure determination of the NSP2 apoprotein and the NSP2-RNP complex. A:** A representative cryo-EM micrograph of the NSP2 apoprotein and corresponding 2D class averages. **B:** The Fourier shell correlation (FSC) curves for the overall NSP2 apoprotein map (orange) and phase randomized map (blue). The map has a final resolution of 3.9 Å. Model versus map FSC curve (green) indicates a model resolution of 4.0 Å. **C:** A representative cryo-EM micrograph of th eNSP2 RNP complex and corresponding 2D class averages. **D:** The Fourier shell correlation (FSC) curves for the overall NSP2-RNP complex map (orange) and phase randomized map (blue). Map has a final resolution of 3.2 Å. Model versus map FSC curve (green) indicates a model resolution of 3.1 Å. **E & G:** Euler angle distribution of particles corresponding to NSP2 apoprotein and NSP2-RNP reconstructions. The cylinder height and colour represent the number of particles (blue to red – low to high). A D4 symmetry has been applied for both the NSP2 apoprotein and the NSP2-RNP complex reconstructions. **F & H:** NSP2 apoprotein (F) and NSP2-RNP complex (H) reconstructions coloured by local resolution as calculated by Relion. **I:** Aligned models of the NSP2 apoprotein and the NSP2-RNP complex built from reconstructions presented in this study (yellow and red, respectively), and a previous crystal structure of the NSP2 apoprotein (PDB ID 1L9V) (green). Alignment of the apoprotein and RNP models from this study to 1L9V had an RMSD of 1.081 Å and 0.772 Å, respectively. The two models from this study had an RMSD of 0.742 Å. **J:** Fitting of previous NSP2 crystal structures into the NSP2 apoprotein cryo-EM density map. The “open” NSP2 conformation (PDB 4G0A) has the CTR flipped outwards, while the “closed” conformation (PDB 1L9V) has the CTR making contacts with the $NSP2_{core}$. The open conformation is not represented by the cryo-EM map, as the CTH density is clearly visible and localized to the $NSP2_{core}$. The green box highlights the cryo-EM density corresponding to the C-terminus.


Autopicking
1,268,983
C2D
955,893
C3D
184,188
(19%)
66,444
(6%)
35,864
(14%)
352,403
(37%)
99,500
(10%)
117,494
(12%)
C2D
305,164
R3D
D4
M+P
4.42 Å
4.34 Å
R3D
D4
Per-particle CTF correction
Particle polishing
M+P
4.34 Å
4.11 Å
C3D (no alignments)
142,357
(47%)
53,416
(17%)
109,391
(36%)
R3D
D4
4.89 Å
R3D
D4
4.79 Å
R3D
D4
4.34 Å
M+P
4.51 Å
M+P
4.26 Å
M+P
3.90 Å (used for
model building)

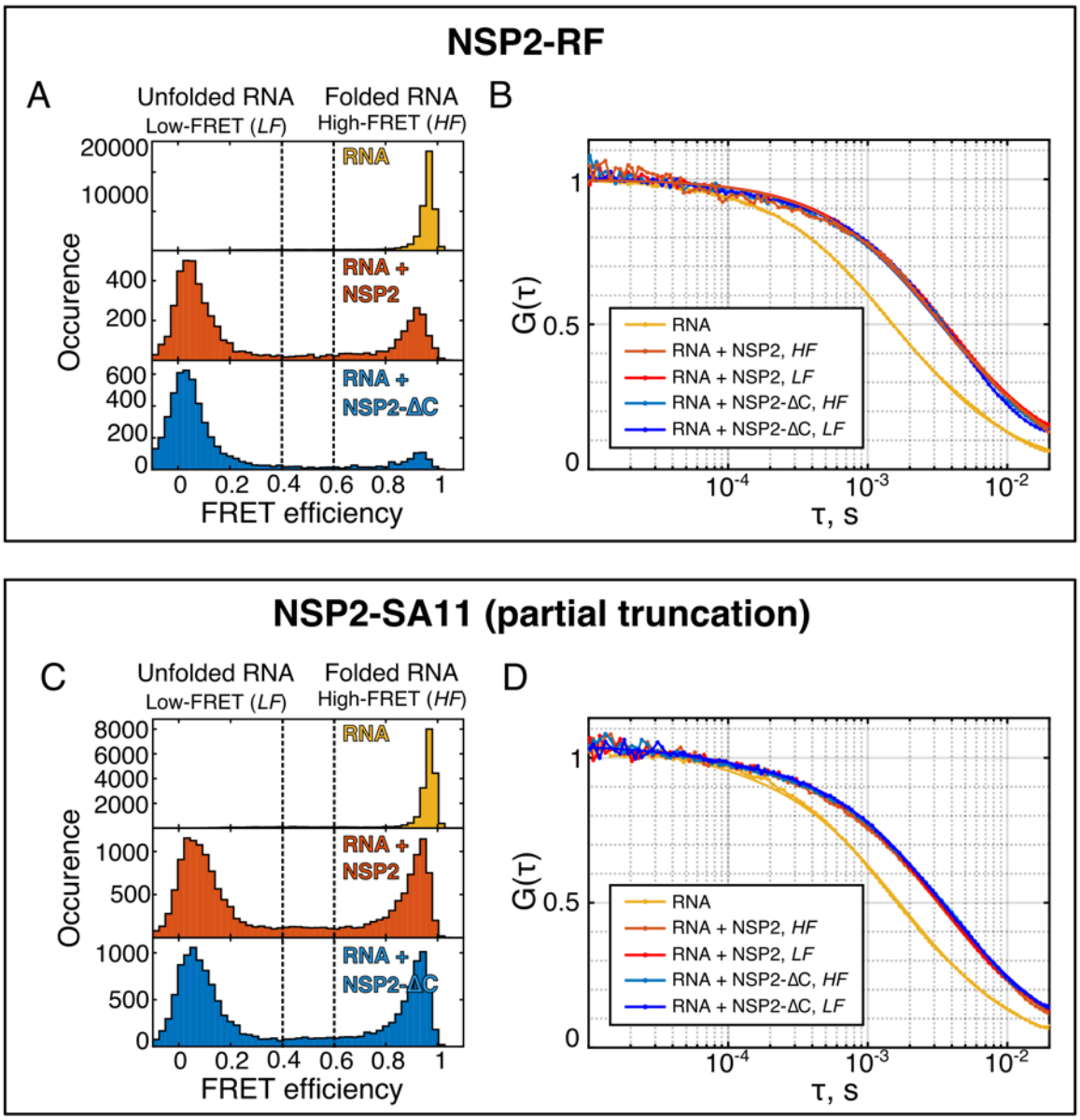
Supplementary Figure S3. The NSP2 apoprotein cryo-EM image-processing workflow. The image processing workflow used to determine the 3D reconstruction of the NSP2 apoprotein. Initial rounds of 2D classification (C2D) and 3D classification (C3D) were performed using Fast Subsets and image alignments with 25 iterations. The highest quality 3D class average was subjected to further 2D classification without fast subsets, and used for 3D refinement (R3D) with D4 symmetry imposed. After per-particle CTF correction and particle polishing, further 3D classification was performed without fast subsets, and without performing image alignment. 3D reconstructions were ultimately subjected to masking and post-processing (M + P).



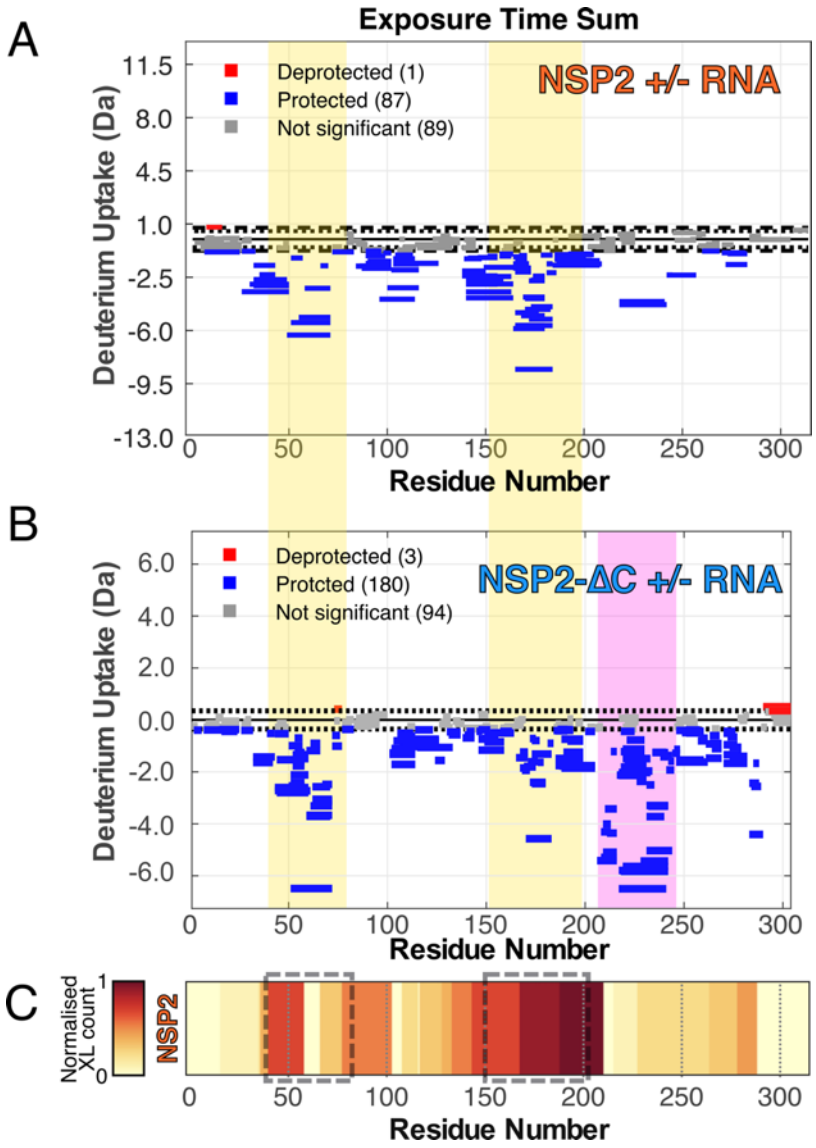
Supplementary Figure S4. Cryo-EM density of the flexible linker within the CTR. **A & B:** Unsharpened (A) and sharpened (B) NSP2 apoprotein cryo-EM density maps. Density corresponding to the CTR is shown in red (green outline). Linker density is diffuse in the sharpened map (B). **C:** Local resolution of NSP2 with the CTR outlined in green.



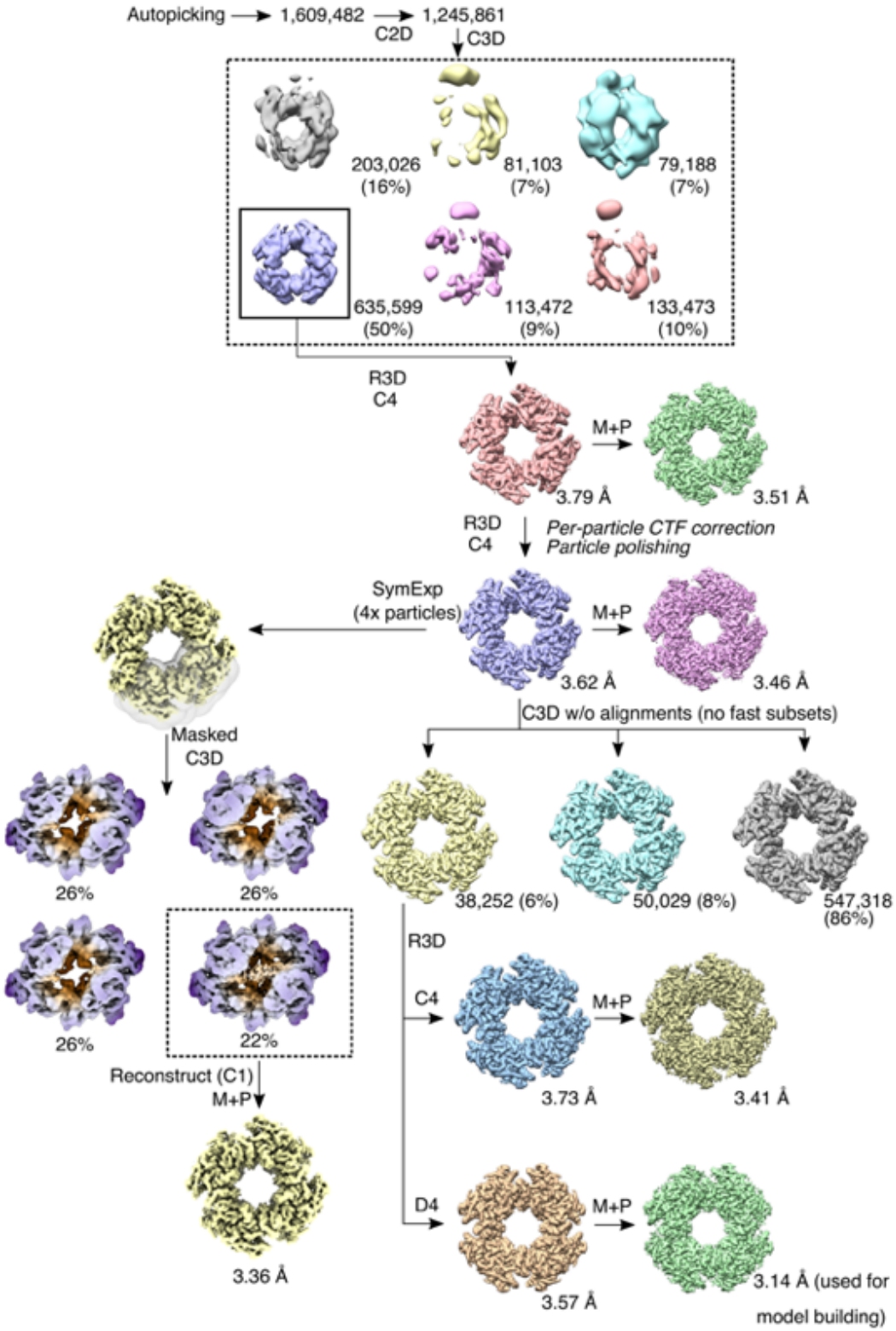
Supplementary Figure S5. Fluorescence cross-correlation spectroscopy (FCCS) and Fluorescence Correlation Spectroscopy (FCS) measurements of NSP2 and NSP2-ΔC. A: Autocorrelation functions (ACFs) of rotavirus (RV) segments S5 and S11 in the presence of NSP2 and NSP2-ΔC (dashed lines). Cross-correlation functions (CCFs) (bold lines) of segments S5 and S11 in the presence of NSP2 (burgundy) and NSP2-ΔC (navy) and segments S6 and S11 in the presence of NSP2 (black) and NSP2-ΔC (magenta). **B:** ACFs and CCFs of S6 and S11 in the presence of NSP2 and NSP2-ΔC.



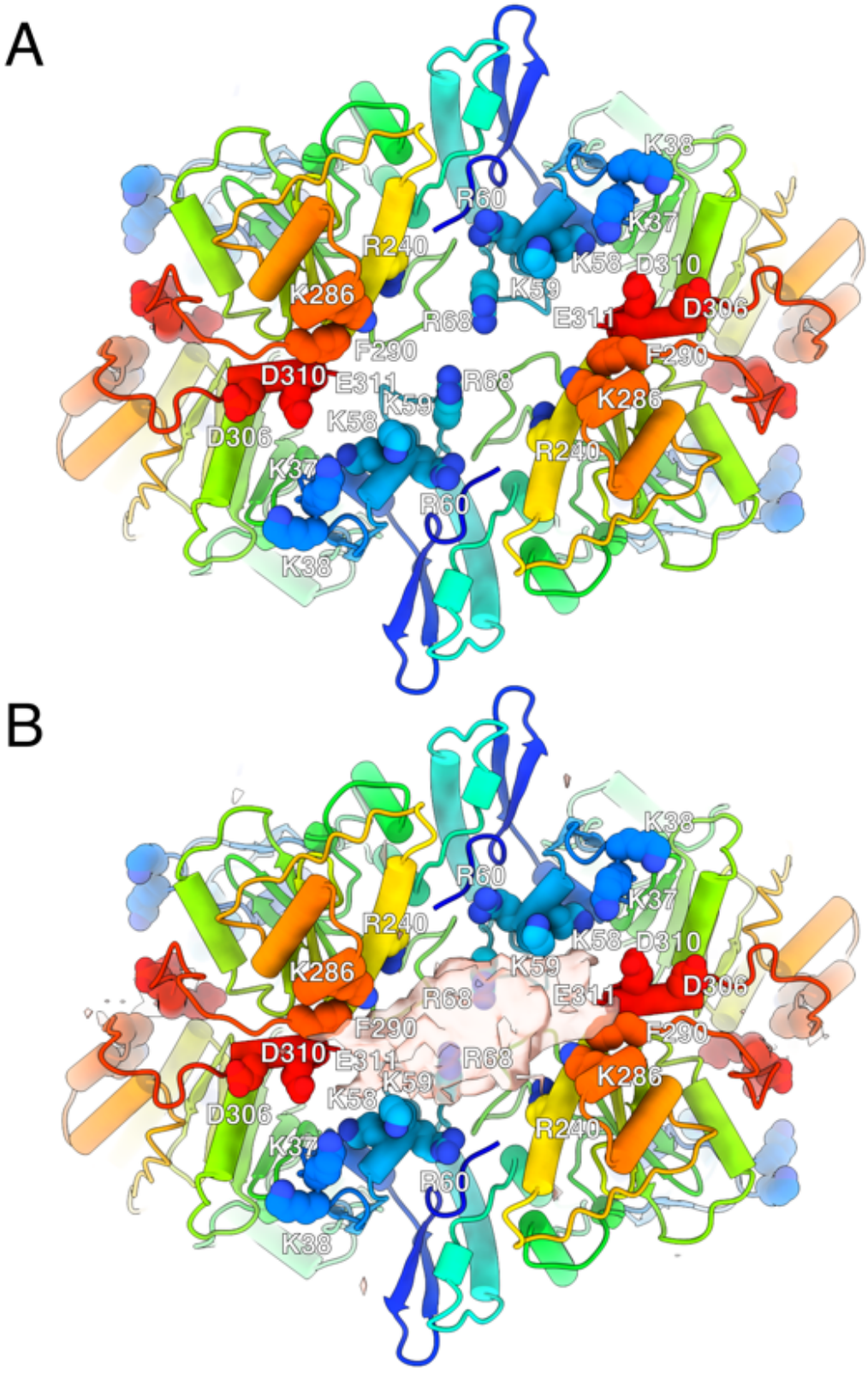
Supplementary Figure S6. SmFRET and burstwise FCS analysis of stem-loop unwinding by NSP2 variants. A & B: SmFRET histograms (**A**) and corresponding burstwise FCS analysis (**B**) of the NSP2 RF strain and NSP2(RF)-ΔC. **C & D** SmFRET histograms (**C**) and corresponding burstwise FCS analysis (**D**) of the NSP2 SA11 strain and a partially truncated NSP2 (lacking the unstructured residues 314-316). These residues are not essential for virus replication (*25*), and do not contribute to RNA unwinding.



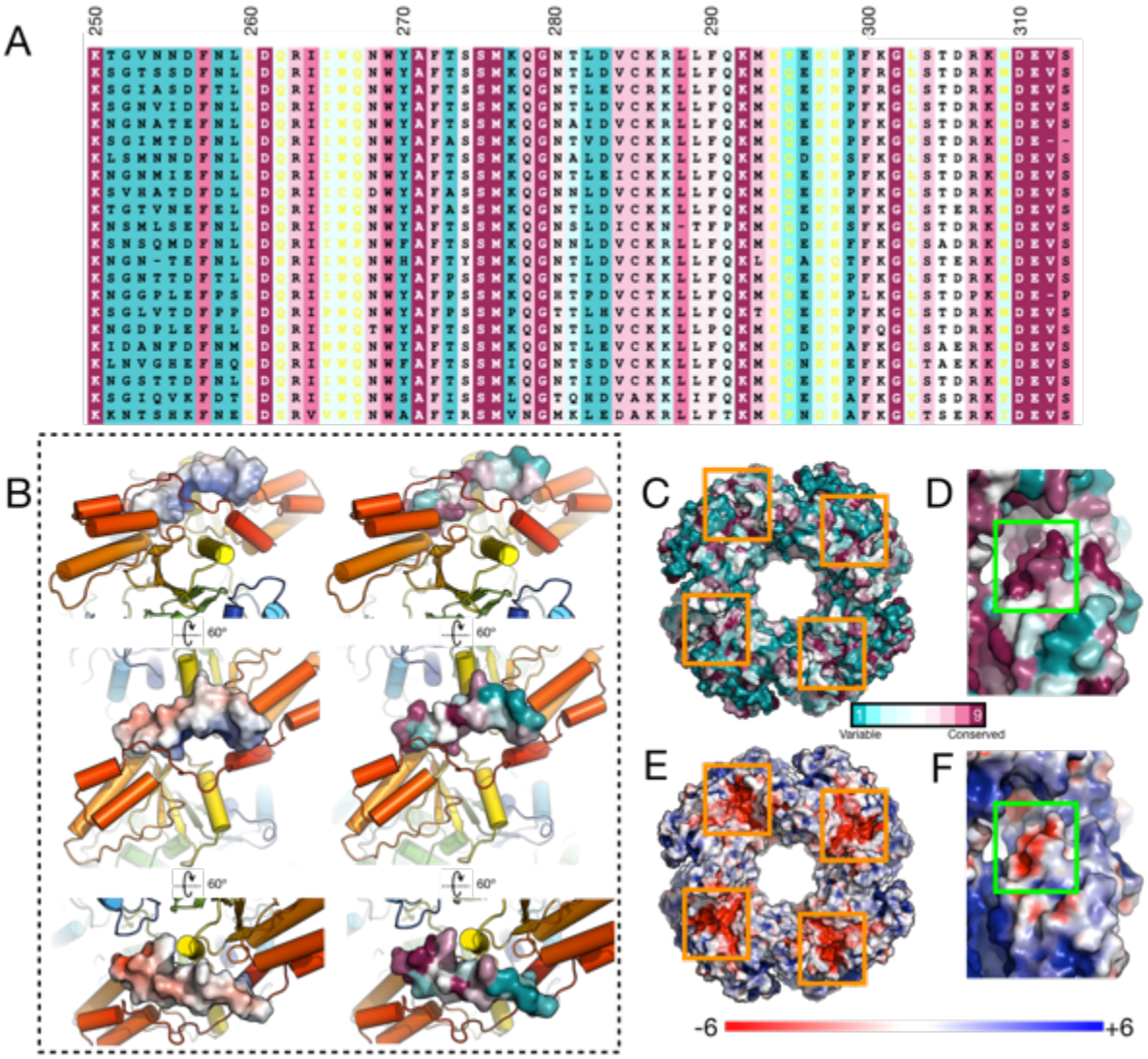
Supplementary Figure S7. A & B: Woods plots showing summed differences in deuterium uptake in NSP2 **(A)** and NSP2-ΔC **(B)** over all timepoints, comparing the apoproteins with their respective RNP complex. Regions corresponding to flexible loops located within the polar, equatorial groove of NSP2 are denoted by yellow shading. In NSP2-ΔC, the magenta box denotes residues that would be otherwise buried underneath the CTR. This includes R240. **C**: A heatmap of crosslinked peptides mapped onto the NSP2 sequence (plotted as counts per residue). Residues corresponding to flexible loops within the equatorial grove are denoted by grey boxes. These regions correspond to the yellow-shaded areas in the above Woods plots.



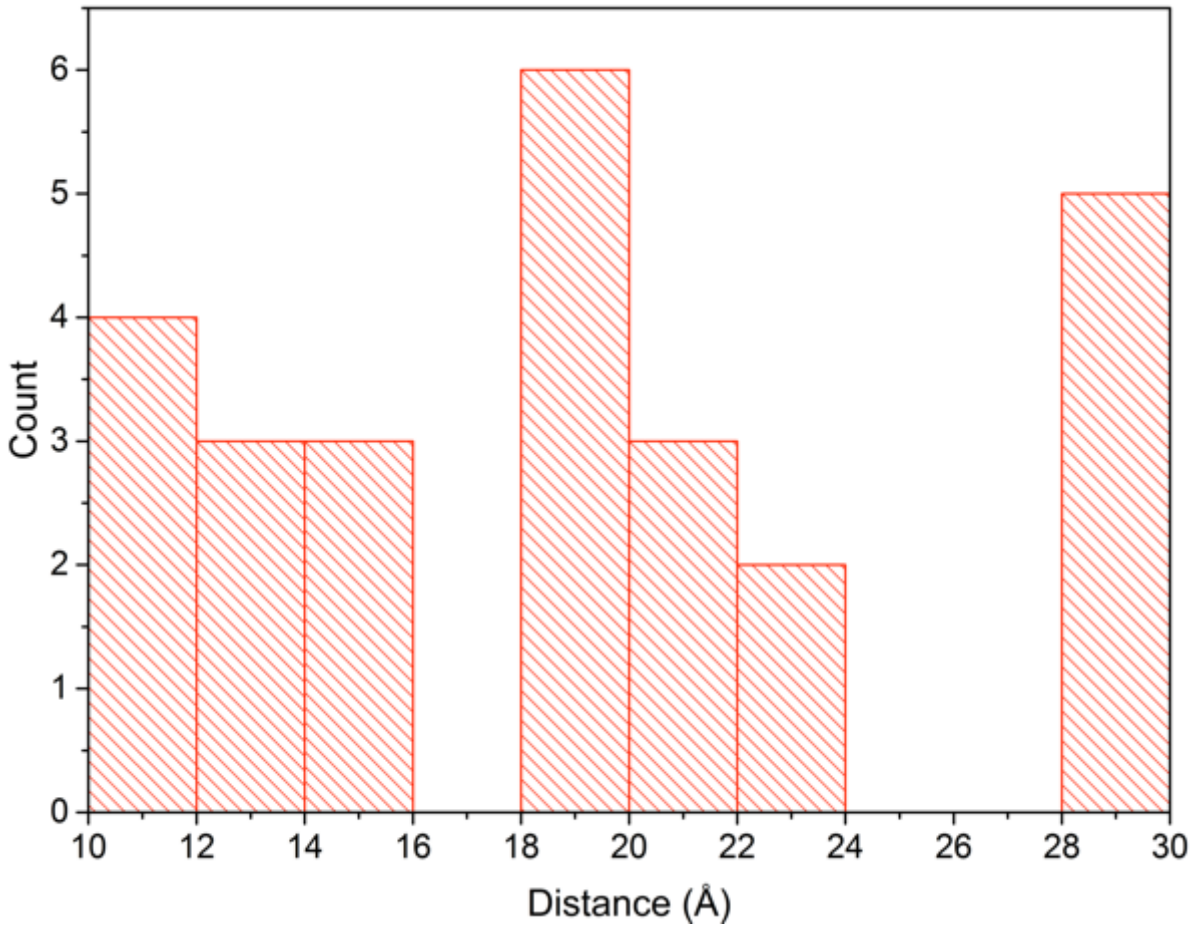
Supplementary Figure S8. The NSP2 RNP cryo-EM image-processing workflow. The image processing scheme used to determine the 3D reconstruction of NSP2 RNP. Initial 2D and 3D classifications (C2D and C3D, respectively) were performed with fast subsets. Purple: Focused classification of NSP2-RNP demonstrate the existence of an RNA-binding groove with variable occupancy. Exemplary 3D class averages of NSP2-RNP with strong (boxed) and weak (all other) RNA density features.



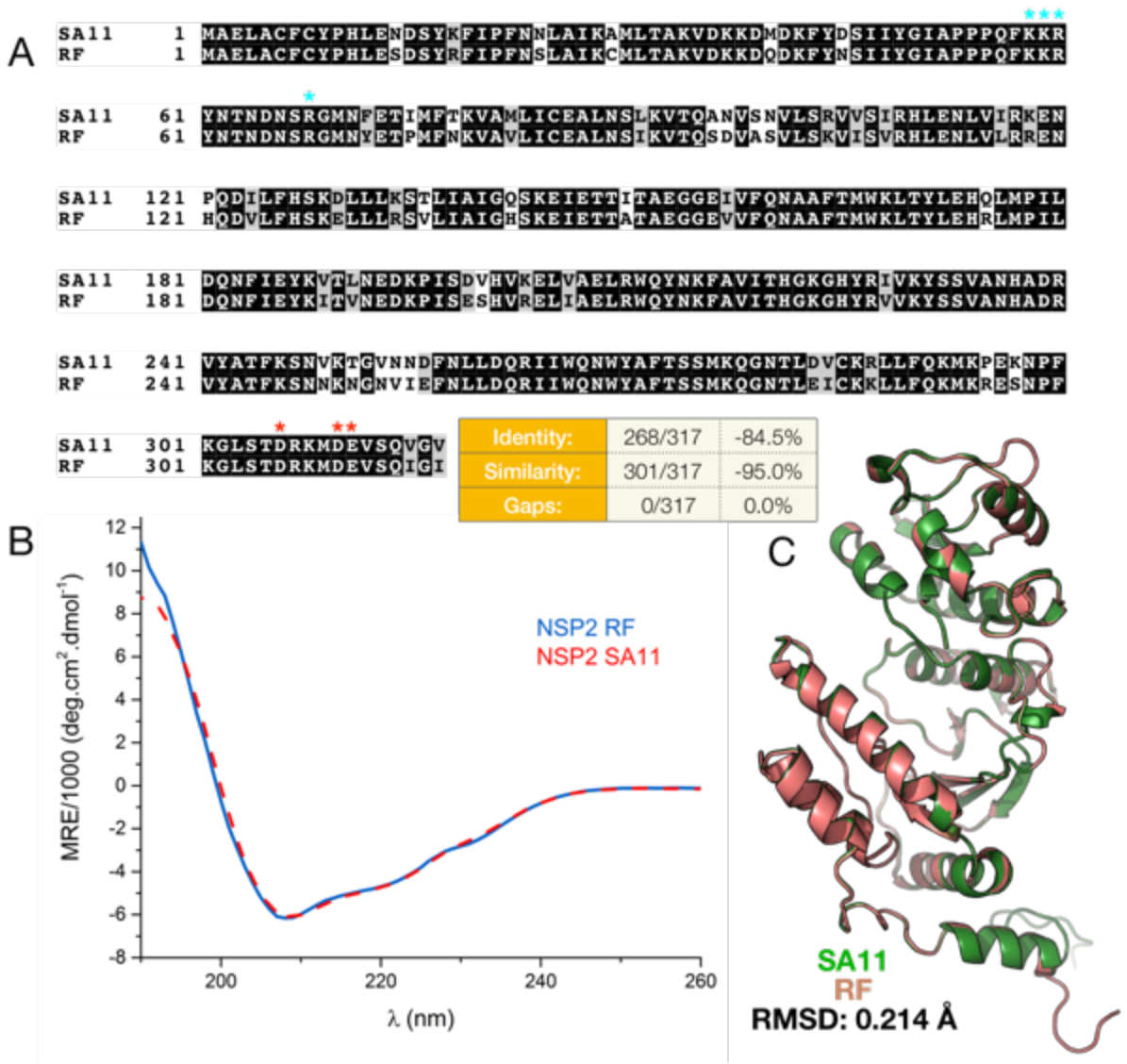
Supplementary Figure S9. Putative RNA-interacting residues. A: An atomic model of NSP2 with potential RNA-interacting residues shown as spheres. **B:** The model shown in A with the positive cryo-EM difference map (i.e. RNA density) superimposed.



Supplementary Figure S10. Conservation & electrostatic analysis of NSP2. A: A ConSurf analysis of the NSP2 C-terminus (residues 250 – 313). **B:** Charge clusters patches are observable on alternate sides of the ampholytic NSP2 CTR surface. The cartoon representation of NSP2 is rainbow coloured according to sequence position. A single CTR is shown as a molecular surface structure. Left: ABPS surface. Right: ConSurf surface conservation surface. **C – F:** Surface-exposed acidic patches within the NSP2 with low (C & E, orange boxes) and high (D & F, green boxes) levels of sequence conservation, as calculated using ConSurf. The highly conserved acidic patch in D & F is within the CTR (D306, D310, E311).



Supplementary Figure S11. Distances between nearest CTR and cryo-EM-identified RNA-binding residues (KKRR K58, K59, R60, R68) and acidic residues within the CTR (DDE D306, D310, E311). For the sake of clarity, only distances between RNA-binding residues and CTR ≤ 30 Å were included. Distances were measured using ChimeraX (*54*).



Supplementary Figure S12. The similarity of NSP2 encoded by the rotavirus strains RF and SA11. A: Alignment of the SA11 and RF NSP2 sequences. RNA-binding residues (as suggested by cryo-EM) are denoted with cyan asterisks. Residues constituting the CTR acidic patch are denoted with red asterisks. Alignment was generated using T-Coffee (*67*) **B**: Circular dichroism (CD) spectra of NSP2 RF (blue) and SA11 (red). **C**: The protein 3D structure prediction of NSP2 RF (pink) was determined using the threading approach using I-TASSER (*68*) based upon the SA11 crystal structure (green). These two atomic models are highly similar, with an overall RMSD between equivalent Cα atoms of the refined model presented here and PDB 1L9V of 0.214 Å.

Supplementary Table S1. Sequences of RNAs used in the study

RNA substrate	Sequence (5' – 3')
Unstructured 20mer (unlabeled) (used for RNA competition assay Fig. 7A)	CUUUUCAAGACAUGCAACAA
Unstructured 20mer (labelled) (used for RNA competition assay Fig. 7A)	AF488-CUUUUCAAGACAUGCAACAA
Unstructured 10mer (labelled) (used for peptide competition assay Fig. 7B)	AF488-CUUCUUUCGA
Biotinylated unstructured 10mer (SPR, Fig. 6)	Biotin-CUUCUUUCGA
Dual-labelled stem loop (smFRET, Fig. 3)	ATTO532-AAAUCUUUGCAAACUAUCCAAUCCAUGCAAAGAUAA-ATTO647N
40mer RNA (cryo-EM)	CUUUUCAAGACAUGCAACAACUUUUCAAGACAUGCAACAA
RV S5 (GenBank ID)	KF729657.1
RV S6 (GenBank ID)	KF729692.1
RV S11 (GenBank ID)	KF729697.1

Supplementary Table S2 Cryo-EM data collection and refinement statistic

	NSP2 apoprotein	NSP2 RNP
Data collection and processing		
Nominal Magnification	75,000	
Voltage (kV)	300 kV	
Defocus range (μm)	-1.2 to -2.4	-1.5 to -3.5
Detector	Falcon III	Gatan K2
Pixel size (Å)	1.065	1.07
Electron exposure ($e^-/Å^2$)	110	55
Symmetry imposed	D4	
Particles in first 3D classification	955,893	1,245,861
Final particle number	109,391	38,252
Nominal map resolution (Å)	3.90	3.14
B-factor*	-239.2	-100.8
FSC threshold	0.143	
Map resolution range (Å)	3.64 – 4.46	3.22 – 4.19
Refinement, model composition and validation		
Initial model used	PDB 1L9V	
Model resolution (Å)	4.0	3.2
FSC threshold	0.5	
Non-hydrogen atoms	20368	20368
RMSD bonds (Å)	0.004	0.008
RMSD angles (º)	0.651	0.810
Ramachandran		
Favoured (%)	92.52	93.01
Allowed (%)	7.48	6.99
Outliers (%)	0.00	0.00
Rotamer outliers (%)	0.00	0.00
Clash score	10.77	7.13
Molprobility score	2.02	1.84
Model-to-map fit		
Cross-correlation coefficient (mask)	0.82	0.83
Cross-correlation coefficient (volume)	0.81	0.81
Main-chain	0.81	0.83
Sidechain	0.78	0.81

Supplementary Table S3 Binding kinetics of NSP2 and NSP2-ΔC as measured by SPR The association (K_{on}) and dissociation (K_{off}) rate constants obtained by SPR are given below.

RNA used	Sample	[NSP2] (nM)	K_{on} ($M^{-1}s^{-1}$)	K_{off} (s^{-1})	K_D (pM)	χ^2	ΔOFF**
10mer RNA	NSP2 WT	6.25	$1.37x10^6$	$9.04x10^{-4}$	6580	1.742	N/A
		12.5	$6.71x10^5$	$8.58x10^{-4}$	1280	15.23	
		25	$5.12x10^5$	$7.20x10^{-4}$	1410	128.7	
	NSP2 ΔC	6.25	$8.24x10^5$	$2.92x10^{-4}$	354	15.85	**3.10**
		12.5	$4.03x10^5$	$2.40x10^{-4}$	596	30.43	**3.58**
		25	$4.95x10^5$	$2.46x10^{-4}$	497	72.05	**2.93**